RICE UNIVERSITY

Decoding biological gene regulatory networks by quantitative modeling

by

Bin Huang

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

DOCTOR OF PHYLOSOPHY

APPROVED, THESIS COMMITTEE

Prof. Jose Onuchic, Chair
Harry C. and Olga K. Wiess Chair of Physics, Professor in Chemistry, and Director, Center for Theoretical Biological Physics, Rice University

Prof. Herbert Levine
Hasselman Professor of Bioengineering, and Director, Center for Theoretical Biological Physics, Rice University

Prof. Cecilia Clementi
Professor of Chemistry and Chemical & Biomolecular Engineering, Rice University

HOUSTON, TEXAS
March 20th, 2017

Decoding biological gene regulatory networks by quantitative modeling

Bin Huang

April 20th, 2017

ABSTRACT

Decoding biological gene regulatory networks by quantitative modeling

by

Bin Huang

Gene regulatory network is essential to regulate the biological functions of cells. With the rapid development of "omics" technologies, the network can be inferred for a certain biological function. However, it still remains a challenge to understand the complex network at a systematic level. In this thesis, we utilized quantitative modeling approaches to study the nonlinear dynamics and the design principles of these biological gene regulatory networks. We aim to explain the existing experimental observations with the model, and further propose reasonable hypothesis for future experimental designs. More importantly, the understanding of the circuit's regulatory mechanism would benefit the design of a *de novo* gene circuit for a new biological function.

We first studied the plasticity of cell migration phenotypes during cancer metastasis, which contains two key cellular plasticity mechanisms - epithelial-to-mesenchymal transition (EMT) and mesenchymal-to-amoeboid transition (MAT). In this study, we quantitatively modeled the core Rac1/RhoA gene regulatory circuit for MAT and later connected it with the core regulatory circuit for EMT. We found four different stable states, consistent with the amoeboid (A), mesenchymal (M), the hybrid amoeboid/mesenchymal (A/M), and the hybrid epithelial/mesenchymal (E/M) phenotypes that are observed in the experiment. We also explored the effects of

microRNAs and EMT-inducing signals like Hepatocyte Growth Factor (HGF), and provided a new insight for the transitions among these phenotypes.

To improve the traditional modeling approaches, we developed a new computational modeling method called Random Circuit Perturbation (RACIPE) to explore the dynamic behaviors of gene regulatory circuits without the requirement of detailed kinetic parameters. We applied RACIPE on several gene circuits, and found the existence of robust gene expression patterns even though the model parameters are wildly perturbed. We also showed the powerful aspect of RACIPE to decipher the operating principles of the circuits.

This kind of quantitative models not only works for gene regulatory network, but also is capable to be extended to study the cell-cell interactions among cancer and immune cells. The results shown the co-occurrence of three cancer states: low risk cancer with intermediate immunity (L), intermediate risk cancer with high immunity (I) and high risk cancer with low immunity state (H). We further used the model to assess the different combinations of cancer therapies.

Acknowledgments

I would like to express my deep appreciation to my advisors, Prof. Jose Onuchic and Prof. Eshel Ben-Jacob, for their patience, inspiration and immense knowledge to support my Ph.D. research. None of the work in the thesis would be possible without their guidance and help. It is my privilege to work with them. Eshel has passed away in 2015, but the memory of him will always be my heart.

In addition, I would like to thank Prof. Herbert Levine, Prof. Peter Wolynes, and Prof. Mingyang Lu, for their insightful comments and hard questions on my projects to incent me to explore my research from various perspectives.

I'm also grateful to my committee member, Prof. Cecilia Clementi, for her help and advice on both my Ph.D. Qualifying Exam and the Thesis Defense.

Many thanks to my group members, including but not limited to Dr. Mohit Jolly Kumar, Dongya Jia, and Dr. Jingcheng Feng, for the sharing of their knowledge and the fantastic collaborations on projects. Also, I thank all CTBP members and my friends at Rice University for the helpful discussions and all the fun we have in the last five years.

Last but not the least, I would like to thank my parents and my girlfriend for their unconditional love and infinite support throughout my Ph.D. research and my entire life.

Contents

List of Figures

Chapter 1

Introduction

1.1. Biological gene regulatory networks

Cell fates are regulated by a substantial number of genes. Instead of acting alone, genes interact with each other and form a very complex gene regulatory network to tightly control cell behaviors[1,2]. This type of collective regulation raises the new challenge to understand the functions of genes at a systematic level, which would be crucial to decipher the association between the genotype and phenotype of cells and further contribute to the synthesis of new circuits for de novo biological functions[3,4].

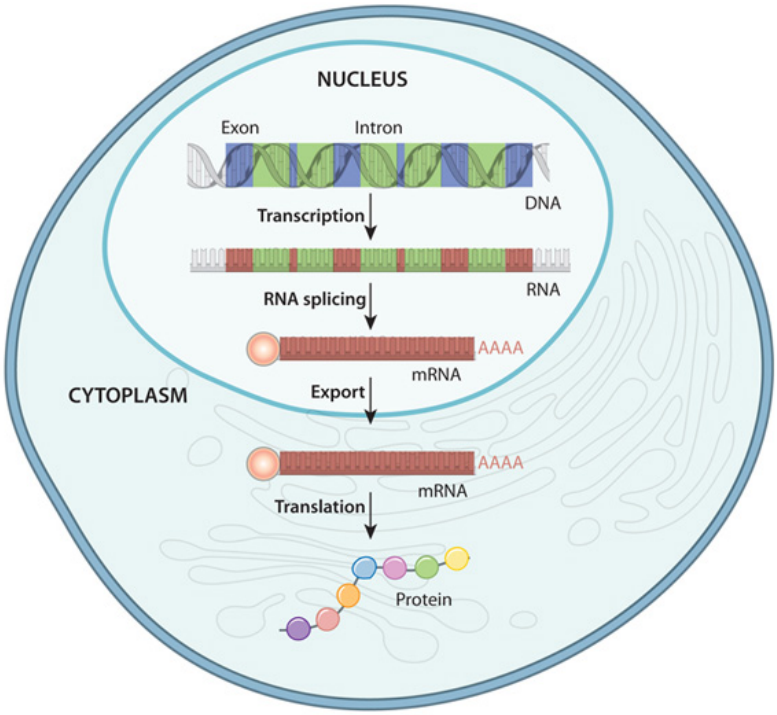


Figure 1.1 Process of gene expression

During transcription, a gene serves as a template to synthesize the precursor mature message RNA (Pre-mRNA) by RNA polymerase. Some regions (Introns) of Pre-mRNA are removed before it is exported out of the nucleus. In cytoplasm, mRNA can be used by the ribosomes to construct a protein dictating certain functions. Adopted from http://www.nature.com/scitable/topicpage/gene-expression-14121669.

Genes are a part of DNA that encode the information of proteins (Figure 1.1). During transcription, a gene serves as a template to synthesize the precursor mature message RNA (Pre-mRNA) by RNA polymerase. Some regions (Introns) of Pre-mRNA are removed before it is exported out of the nucleus. In cytoplasm, mRNA can be used by the ribosomes to construct a protein dictating certain functions. Several steps in the gene expression process can be regulated. For example, epigenetic regulations could chemically change DNA methylation and histone modifications to silent or activate the expression of genes[5]. Also, the protein expressed by one gene may be able to move

back to the nucleus to bind to the promoter regions of the other genes to transcriptionally inhibit or activation the genes' expressions. These multiple layers of regulations increase the difficulty in understanding the mechanism of gene regulatory network.

As a possible solution to these challenges, system biology has been rapidly developed in the past few decades to provide a system-level understanding of the complex biological behaviors[6,7]. There are two general approaches for system biology: "top-down" approach to integrate the omics data to reconstruct the regulatory network to discover new insights of biological behaviors, and "Bottom-up" approach to model detail interactions among molecules to explain the operational mechanism behind certain biological behaviors. These two approaches have been successfully applied in several areas. For instance, "Top-down" approaches are used to reconstruct the metabolic networks[8,9], and drug discovery for diseases[10]. Also, "Bottom-up" approaches have been used for the modeling of biological process such as Epithelial-to-Mesenchymal transition (EMT) during cancer metastasis[11–13], cell cycles[14,15], and circuit design in synthetic biology, such as genetic toggle switch[3], and repressilator[4]. However, these two approaches have largely been developed separately until now. Recently, several works[16] including our work in Chapter 3 and 4 start to bridge these approaches as a more powerful method to understand complex biological problems.

1.2. Quantitative modeling approaches

This thesis mainly focuses on using "Bottom-up" approaches to study gene regulatory networks. In "Bottom-up" approaches, we assumed that for a certain

biological function to study, there exists a core gene circuit containing the master regulatory genes for the function. The other genes outside the core circuit have little effects on this specific function. The core gene circuit can be constructed according to either the literature, interaction databases (e.g. Ingenuity pathway analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity), KEGG[17], GO[18]), or computational methods[19].

There are typically two ways to study the core gene circuit in "Bottom-up" approaches, one is by the deterministic model, such as ordinary differential equations (ODEs), partial differential equations (PDEs), and the other is by stochastic model, such as master equations or Langevin equations. The work in this thesis mainly used the deterministic model to study the nonlinear dynamics and time dynamics of the core gene circuits. Here, I will take a simple gene circuit, toggle switch, to illustrate the modeling approach and the analysis tools.

As shown in Figure 1.2a, the toggle switch circuit is two genes (A and B) mutually inhibiting the expression of each other. When gene A is expressed, protein A could move back into nucleus and bind the promoter region of gene B to inhibit its expression. Gene B could also inhibit the expression of gene A in a similar way. The chemical reactions of these processes are described as following:

$$\text{Gene A} \xrightarrow{g_{a0}} \text{mRNA A} \quad (1)$$

$$\text{Gene A} + \text{Protein B} \underset{k_{dB}}{\overset{k_{aB}}{\rightleftarrows}} \text{Gene A} - \text{Protein B} \quad (2)$$

$$\text{Gene A} - \text{Protein B} \xrightarrow{g_{a}} \text{mRNA A} \quad (3)$$

$$\text{mRNA A} \xrightarrow{g_{am}} \text{Protein A} \quad (4)$$

$$\text{mRNA A} \xrightarrow{K_{mA}} \varnothing \quad (5)$$

$$\text{Protein A} \xrightarrow{K_{A}} \varnothing \quad (6)$$

$$\text{Gene B} \xrightarrow{g_{b0}} \text{mRNA B} \quad (7)$$

$$\text{Gene B} + \text{Protein A} \underset{k_{dA}}{\overset{k_{aA}}{\rightleftarrows}} \text{Gene B} - \text{Protein A} \quad (8)$$

$$\text{Gene B} - \text{Protein A} \xrightarrow{g_{b}} \text{mRNA B} \quad (9)$$

$$\text{mRNA B} \xrightarrow{g_{bm}} \text{Protein B} \quad (10)$$

$$\text{mRNA B} \xrightarrow{K_{mB}} \varnothing \quad (11)$$

$$\text{Protein B} \xrightarrow{K_{B}} \varnothing \quad (12)$$

where g_{a0} is the basal transcriptional rate for gene A, while g_a is the inhibited transcriptional rate when protein B binds to the promoter region of gene A. k_{aB} and k_{dB} are the binding and unbinding rate for protein B to the promoter region of gene A. g_{am} is the translational rate for mRNA A, and K_m and K_A are the degradation of mRNA A and protein A. Similar meanings for the kinetic parameters for gene B. These 12 chemical reactions can be modeled by the set of six ordinary differential equations (ODEs) as following:

$$\begin{aligned}
\frac{d_{GAB}}{dt} &= k_{aB} \bullet GA \bullet B - k_{dB} \bullet GAB \\
\frac{d_{mA}}{dt} &= g_{a0} \bullet GA + g_a \bullet GAB - K_{mA} \bullet mA \\
\frac{dA}{dt} &= g_{am} \bullet mA - K_A \bullet A \\
\frac{d_{GBA}}{dt} &= k_{aA} \bullet GB \bullet A - k_{dA} \bullet GBA \\
\frac{d_{mB}}{dt} &= g_{b0} \bullet GB + g_b \bullet GBA - K_{mB} \bullet mB \\
\frac{dB}{dt} &= g_{bm} \bullet mB - K_B \bullet B
\end{aligned} \quad , (1)$$

where GAB, mA and A are the concentration of gene A-protein B complex, the mRNA A and the protein A respectively. Usually, the binding/unbinding process of regulators (protein A and B) to the promoter regions are much faster than the transcription and translation processes, thus we can assume that these processes get to the equilibrium very fast, namely $d_{GAB}/dt = 0$ and $d_{GBA}/dt = 0$ in Equation (1), and after combining it with the principle of mass balance, we can get

$$\begin{aligned}
GAB &= \frac{B}{B + k_{dB}/k_{aB}} \bullet GA_{tot} = H^+(B) \bullet GA_{tot} \\
GA &= \frac{k_{dB}/k_{aB}}{B + k_{dB}/k_{aB}} \bullet GA_{tot} = H^-(B) \bullet GA_{tot} \\
GBA &= \frac{A}{A + k_{dA}/k_{aA}} \bullet GB_{tot} = H^+(A) \bullet GB_{tot} \\
GB &= \frac{k_{dA}/k_{aA}}{A + k_{dA}/k_{aA}} \bullet GB_{tot} = H^-(A) \bullet GB_{tot}
\end{aligned} \quad , (2)$$

where GA_{tot} is the total concentration of gene A, while GB_{tot} is the total concentration of gene B. Also, the degradation of mRNA is typically four times faster than that of protein[20], so we can further assume that the mRNAs achieve the equilibrium fast,

namely $d_{mA}/dt = 0$ and $d_{mB}/dt = 0$ in Equation (1). Therefore, we can further reduce the model to two ODEs to describe the behavior of toggle switch circuit:

$$\begin{aligned}
\frac{dA}{dt} &= g_{a0}^{'} \bullet H^{-}(B) + g_{a}^{'} \bullet H^{+}(B) - K_A \bullet A \\
\frac{dB}{dt} &= g_{b0}^{'} \bullet H^{-}(A) + g_{b}^{'} \bullet H^{+}(A) - K_B \bullet B \\
g_{a0}^{'} &= g_{a0} \bullet g_{am} \bullet GA_{tot} / K_{mA} \\
g_{a}^{'} &= g_{a} \bullet g_{am} \bullet GA_{tot} / K_{mA} \\
g_{b0}^{'} &= g_{b0} \bullet g_{bm} \bullet GB_{tot} / K_{mB} \\
g_{b}^{'} &= g_{b} \bullet g_{bm} \bullet GB_{tot} / K_{mB}
\end{aligned} \quad , (3)$$

Typically, we can use numerical simulations to analyze the temporal or nonlinear dynamics of the model. By using the Euler method or Runge-Kutta methods (e.g. ODE solver in Matlab), we can solve the ODEs to obtain the temporal trajectories of genes in the circuit starting from a certain initial condition. Notably, for a multi-stable system such as toggle-switch circuit, different initial conditions will finally end up at different stable steady states after a long-time simulation (Figure 1.2b).

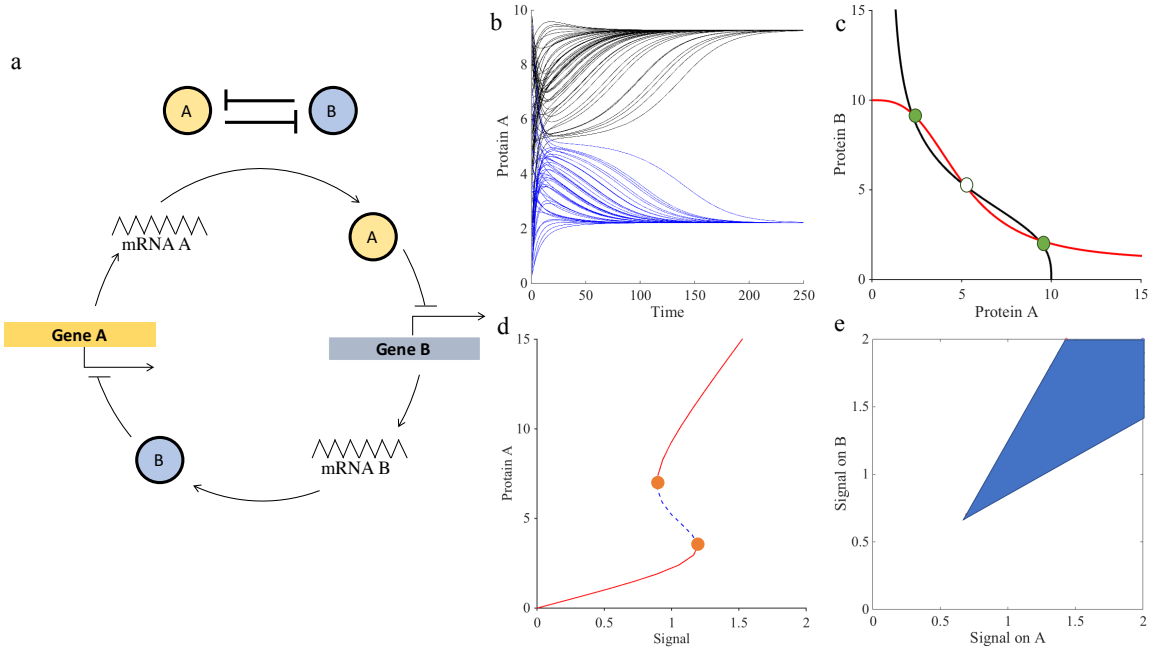


Figure 1.2 Modeling and analysis of toggle switch circuit.

(a) Details about toggle switch circuit. Gene A and gene B mutually inhibit the expression of each other. The proteins expressed by one gene are able to get into the nucleus and transcriptionally inhibit the expression of the other gene by binding to the promoter region of that gene. **(b)** Temporal trajectories of toggle switch circuit starting from 100 different initial conditions. There are two different stable states of the system. **(c)** Nullclines for toggle switch circuit, where black nullcline is for $dA/dt = 0$ and red nullcline is for $dB/dt = 0$. Green solid circles denote the stable fixed points, and green hollow circles denote the unstable fixed points. **(d)** One parameter bifurcation of toggle switch circuit when gene A is regulated by the external signal (I). In Equation (3), we multiply the g'_{a0} and g'_a by the level of signal (I). Red lines stand for stable states while the blue line stands for the unstable states. The orange dots represent the transition (bifurcation) points. **(e)** Two parameters bifurcation when gene A and gene B are both

regulated by the external signals. In Equation (3), we multiply the $g_{a0}^{'}$ and $g_{a}^{'}$ by the level of signal I_A, and multiply the $g_{b0}^{'}$ and $g_{b}^{'}$ by the level of signal I_B. Blue area stands for the bi-stability while the white area stands for the mono-stability.

To thoroughly explore all possible stable steady states of a model, there are usually two different methods. First method is an extension of above simulations. Instead of running the simulation from one initial condition, we can repeat the long-time simulation from thousands of different initial conditions and identify the distinct solutions from all the solutions. As shown in Figure 1.2b, we used this method with 100 different initial conditions to find the two stable steady states for the toggle switch circuit. This method can be applied to very large gene regulatory network, but may miss out some rare stable states. For second method, we can calculate the nullclines to find out the stable steady states. To use this method, the model should be able to be reduced to two ODEs with two variables (x and y) by setting all other ODEs to be zeros and representing the other variables as the function of x and/or y.

$$\begin{aligned} \frac{dx}{dt} &= f(x,y) \\ \frac{dy}{dt} &= g(x,y) \end{aligned}, \quad (4)$$

A x-nullcline is defined as a set of points on the x-y dimensions so that dx/dt = 0. Similarly, a y-nullcline is a set of points so that dy/dt = 0. The intersections of these two nullclines are the steady states for the whole system. The stability of the steady states can be further determined by the linear approximation method[21]. This method can explore all possible stable steady states, but the restriction of this method is obvious since it

required the possible reduction of the ODEs model. Therefore, this method only works well for a small gene circuit, such as toggle switch circuit (Figure 1.2c). This limit of the method could be released by using continuation method[22] to calculate the nullclines for a high-dimensional model, which are provided by MATCONT[23] package.

Furthermore, we can apply bifurcation analysis to understand the response of circuit stability against the modification of model parameters, which may be caused by some external signals. The bifurcation diagrams can be calculated either by the nullcline method at different level of signals or by the continuation method provided by MATCONT[23]. Taking toggle switch circuit as an example, if gene A receives a constant transcriptional activation by an activator, one-parameter bifurcation diagram could show the circuit stability at different level of the signal and indicate the transitions between different stable states (Figure 1.2d). We can also calculate two-parameter bifurcation phase diagram if both gene A and gene B receive external signals (Figure 1.2e).

1.3. Scope of the thesis

This thesis aims to use quantitative modeling approaches to understand the important roles of gene regulatory circuits in controlling biological functions. The remainder of the thesis is organized as follow:

- Chapter 2 focuses on the understanding of the plastic cell migration phenotypes during cancer metastasis via quantitatively modeling the involved core gene circuit.

- Chapter 3 presents the development of a new computational method, Random Circuit Perturbation (RACIPE), unbiasedly model a gene circuit with the absent of accurate model parameters. We applied RACIPE on several simple circuits and a biological circuit controlling epithelial-to-mesenchymal transition of cancer cells.
- Chapter 4 shows the application of RACIPE method on stem cell circuit, which controls the early development of mouse embryo. The predictions by RACIPE have good agreements with single cell genomic data.
- Chapter 5 extended the traditional modeling approached for gene regulatory network towards the interaction network among cancer-immune cells. We used the model for the assessment of different therapeutic treatments.

Chapter 2

Towards elucidating the plasticity of cell migration phenotypes during cancer metastasis

Cellular plasticity during cancer metastasis is a major clinical challenge. Two key cellular plasticity mechanisms - epithelial-to-mesenchymal transition (EMT) and mesenchymal-to-amoeboid transition (MAT) – have been widely investigated, yet a comprehensive understanding of their underlying mechanisms remain enigmatic. These transitions are mediated by two core gene regulatory circuits - miR-200/ZEB/miR-34/SNAIL circuit for EMT and Rac1/RhoA circuit for MAT. The regulatory mechanism of the EMT circuit is well studied in the previous work of our lab[24]. Here, we focused on the study of Rac1/RhoA circuit for MAT by using detailed modeling of GTPase-based regulations. We found that the circuit can operate as a three-way switch. We proposed to associate the circuit's three possible states to the amoeboid (A), mesenchymal (M) and amoeboid/mesenchymal (A/M) hybrid phenotype. Furthermore, we studied the coupling

between these two core circuits by considering the two microRNAs (miR-200 and miR-34) as external signals to the core MAT circuit. We showed that this coupled circuit enables an additional stable steady state that corresponds to the hybrid epithelial/mesenchymal (E/M) phenotype. Our model also recapitulated the metastasis-suppressing role of the microRNAs even in the presence of EMT-inducing signals like Hepatocyte Growth Factor (HGF). It also enabled mapping the microRNA levels to the transitions among various cell migration phenotypes. These results offer a mechanistic understanding for the observed phenotypic transitions among different cell migration phenotypes during cancer metastasis.

2.1. Introduction

Plasticity of cell migration phenotypes during cancer metastasis

Metastasis causes more than 90% of cancer-related deaths[25,26]. For carcinomas, the most common type of tumors, metastasis begins when some epithelial cells from the primary tumor lose their apico-basal polarity and cell-cell adhesion and acquire migratory and invasive characteristics, through a process known as Epithelial–to-Mesenchymal Transition (EMT)[27]. Cells can undergo a partial or complete EMT and consequently move collectively or individually while treading through the extra-cellular matrix (ECM) and circulating in the bloodstream[28,29]. Upon reaching the secondary site, these circulating tumor cells (CTCs) exit the bloodstream and usually undergo a Mesenchymal-to-Epithelial Transition (MET) to seed metastases[27].

The collectively migrating cells display both epithelial (E) (cell-cell adhesion) and mesenchymal (M) (migration) properties, thereby reflective of the hybrid epithelial/mesenchymal (E/M) or partial EMT phenotype; while the individually moving cells display at least two distinct phenotypes—mesenchymal (M) and amoeboid (A) (here we specifically refer to blebby amoeboid (BA) (Figure 2.1)). Cells of the M phenotype are elongated and spindle-shaped with their leading edge characterized by lamellopodia (LAM) and/or filopodia (FIL). These cells are able to remodel and even degrade the ECM by secreting Matrix Metalloproteinases (MMPs) and therefore act as 'path generators'[29,30]. Conversely, rather than secreting MMPs to remodel the ECM, cells that exhibit the A phenotype have higher shape plasticity that enables them to squeeze into the gaps in the ECM, thus acting as 'path finders'[29,30]. In 3D environment, many carcinoma cells exhibit Mesenchymal-to-Amoeboid Transition (MAT) and Amoeboid-to-Mesenchymal Transition (AMT)[31] either spontaneously or in response to external signals from the microenvironment[32–34]. Recent studies also have identified several individually migratory phenotypes displaying mixed amoeboid and mesenchymal characteristics[31,35–37], indicative of a hybrid amoeboid/mesenchymal (A/M) phenotype[38]. Such rich plasticity allows cancer cells to adapt to the changing microenvironment quickly and facilitates tumor metastasis[27,29,39].

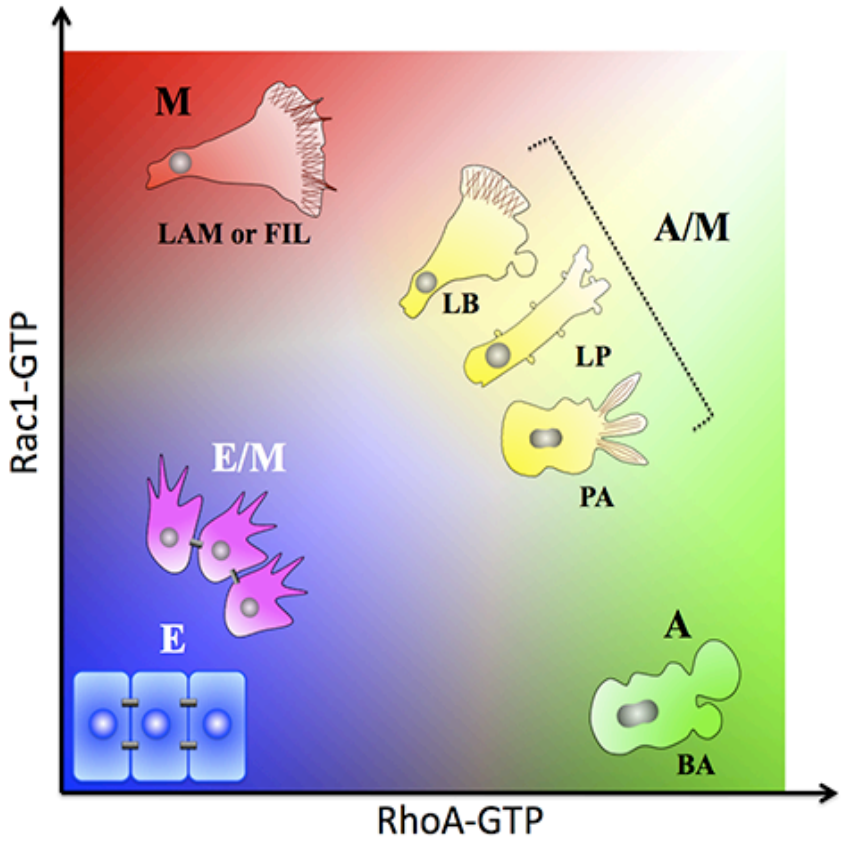


Figure 2.1 Plasticity of cell migration phenotypes in terms of the activities of Rac1/RhoA GTPases.

Background colors correspond to the different level of Rac1-GTP and RhoA-GTP - red for high level of Rac1-GTP, green for high level of RhoA-GTP, yellow for high levels of both of them, and blue for low levels of both of them. The activity of the GTPases is hypothetically associated with the cell morphology and mobility. The phenotypes are depicted as cartoons displaying their corresponding morphological features - epithelial phenotype (E), hybrid epithelial-mesenchymal phenotype (E/M), mesenchymal phenotype (M), amoeboid phenotype (A), and hybrid amoeboid-mesenchymal phenotype (A/M). M phenotype is characterized with lamellopodia (LAM) and/or filopodia (FIL), and A phenotype here is specifically referred to Blebby amoeboid (BA) phenotype, which is characterized by blebbing. The A/M phenotype is considered as a set of different morphologies - Lamellipoida with blebs (LB), Lobopodia (LP) and Pseudopodal amoeboid (PA). The blue color is also associated with strong cell-cell adhesion, as

observed in E or E/M phenotypes, while the rest colors are associated with single cell migration modes.

Core gene regulatory circuits controlling the transition among different migration phenotypes

The choice among the aforementioned phenotypes is operated by the regulations of key genes in the context of complex interactions among them. Our previous theoretical work has explained how the core EMT/MET regulatory circuit allows transitions between E/M phenotype displaying collective cell migration and the mesenchymal (M) phenotype displaying individual migration[40]. The core regulatory circuit consists of two interconnected mutually inhibitory circuits between a microRNA and a transcription factor (TF) – miR-34/SNAIL and miR-200/ZEB4[28] (Figure 2.2b). miR-34/SNAIL acts as an integrator of various external signals for inducing or inhibiting EMT, and feeds to miR-200/ZEB that acts as the three-way decision making switch for EMT/MET, thereby allowing for three distinct phenotypes – E (high miR-200, low ZEB), M (low miR-200, high ZEB) and E/M (medium miR-200, medium ZEB)[40].

Amoeboid-to-Mesenchymal Transition (AMT) is operated by the Rac1/RhoA regulatory circuit. Rac1 and RhoA belong to the Rho family of small GTPases and act as molecular switches by changing between their active (the GTP-bound state) and inactive (the GDP-bound state and the GDI-bound state) forms[41]. This switching process is regulated by three sets of proteins: GEFs (Guanine Nucleotide Exchange Factors) that catalyze the exchange of bound GDP for GTP, thus elevating the levels of the active GTPases; GAPs (GTPase-activating proteins) that promote the intrinsic GTP hydrolysis

rates, thus reducing the concentration of the active GTPases; and GDIs (Guanine nucleotide dissociation inhibitors) that sequester GTPases from the membrane to the cytosol and stabilize the proteins by preventing degradation[42].

Rac1 and RhoA regulate the phenotypic transitions by controlling the actin polymerization and actomyosin contraction[43]. Therefore, the activities of these two GTPases have been observed to correlate with cell morphology and motility. For example, actomyosin contractility increases in response to the RhoA activation, thus resulting in membrane blebbing and facilitating the amoeboid phenotype (A)[35,44,45]. On the other hand, the activation of Rac1 results in the formation of focal adhesions and actin polymerization, which leads to the formation of lamellopodia and enables a mesenchymal phenotype (M)[36,44,46]. Appropriate changes in the relative strengths of these two driving forces (actomyosin contraction vs. actin polymerization) allow for not only the transitions between the A and M phenotypes, but also might enable the transition into/from the hybrid A/M phenotype[31]. These transitions can be triggered by extracellular signals such as Hepatocyte Growth Factor/Scatter Factor (HGF/SF) (Figure 2.2a) through the c-MET pathway[47]. c-MET, the specific tyrosine kinase receptor for HGF/SF, is often overexpressed in many carcinomas and correlates with poor patient survival[48,49]. The activated c-MET recruits Grb2 (Growth-factor receptor bound protein 2) and Gab1 (Grb2 associated binding protein 1) to regulate the activity of both RhoA and Rac1 (Table S1.1 and Figure S1.1). It is contradictory that the HGF/SF/c-MET pathway has been reported to be able to induce either mesenchymal[50,51] or amoeboid[52] phenotype, therefore the underlying mechanisms of the HGF/SF/c-MET pathway remain elusive.

More importantly, these two regulatory circuits are coupled through the microRNAs, miR-200 and miR-34, that inhibit both RhoA as well as Rac1[53–58] (Figure 2.2b). Therefore, the analysis of the coupled circuit between them may help us understand the operating principles of inter-conversion between collective (E/M) and these different individual modes of migration (A, A/M, M).

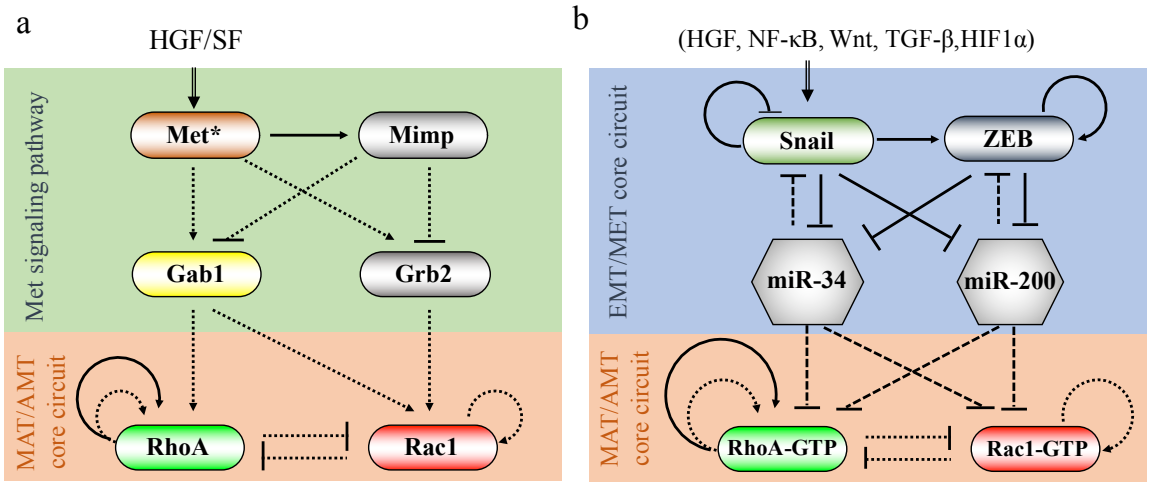


Figure 2.2 Core gene regulatory circuits for MAT and EMT.

(a) Rac1/RhoA gene regulatory circuit connected with the c-MET pathway. The AMT/MAT is mainly regulated by the Rac1/RhoA regulatory circuit, while RhoA and Rac1 are regulated via Grb2 and Gab1 by the c-MET pathway, which receives the external signal from HGF/SF. **(b)** Coupling between EMT and AMT circuit. The core EMT/MET regulatory circuit consists of two coupled mutually inhibitory circuits (SNAIL/miR-34 and ZEB/miR-200). It can receive external EMT-inducing signals such as HGF, and regulate the Rac1/RhoA circuit by inhibiting the translation of RhoA and Rac1 via miR-34 and miR-200. A solid arrow denotes activation, and a solid bar denotes repression. A solid line represents transcriptional regulation, a dotted line represents indirect regulations such as protein-protein interactions, and a dashed line represents translational inhibitions by microRNAs.

The aim of this study was: 1) to elucidate how Rac1/RhoA circuit regulates the AMT/MAT transitions by developing a new framework to model the small GTPase-based regulation in the context of the Rac1/RhoA circuit. 2) to understand the coupling between EMT/MET and AMT/MAT transitions by analyzing the coupling of their core regulatory circuits.

2.2. Methods

Theoretical framework to model the Small GTPase-based Regulatory Circuit.

The challenge posed in modeling the small GTPase-based regulatory circuit is to incorporate the elaborate transitions between different forms of the GTPases. Typically, a GTPase protein can switch among its active (GTP-bound state) and inactive (GDP-bound state and GDI-bound state) forms under the regulation of three sets of proteins (GEFs, GAPs and GDIs)[41]. To understand these features of the GTPases, we developed the theoretical framework for GTPase-based circuit by specifically modeling the detailed molecular interactions as illustrated in Figure 2.3.

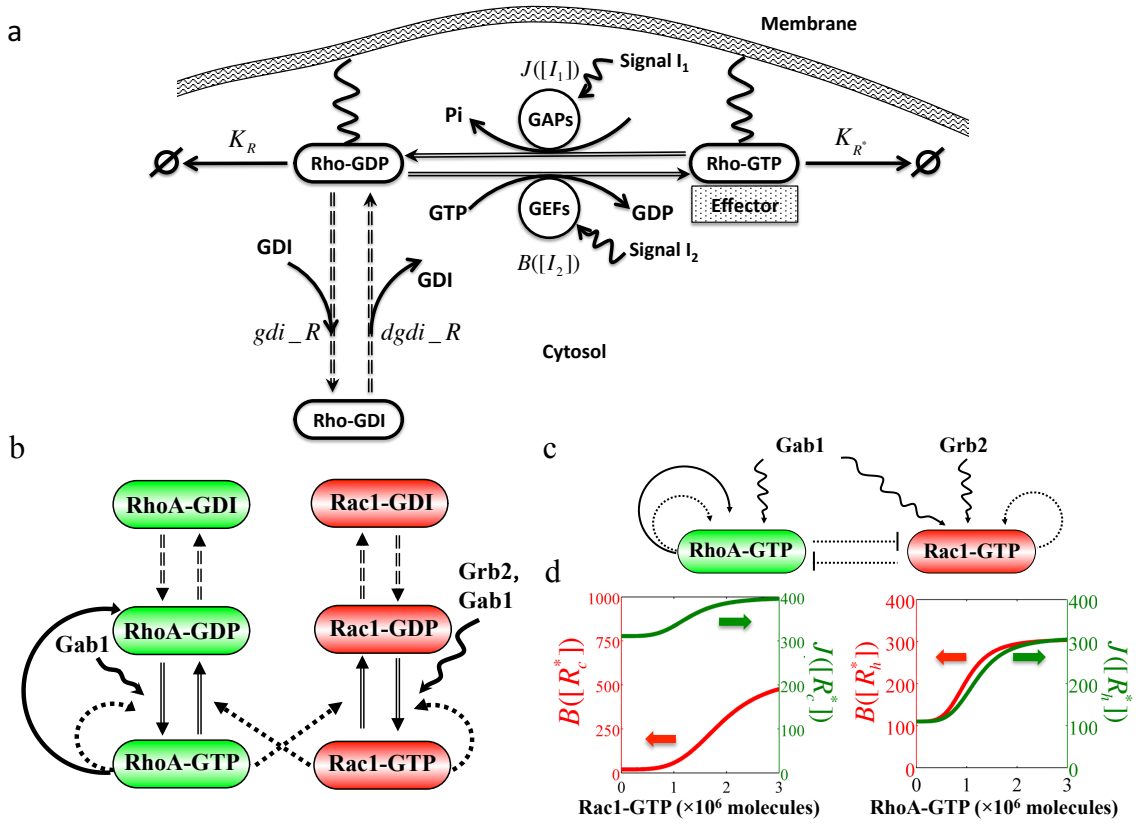


Figure 2.3 Schematic diagram of the regulation of a typical Rho family GTPase (denoted as Rho).

(a) Schematic diagram of the regulation of a typical Rho family GTPase (denoted as Rho). The inactive GDP-bound state of Rho (Rho-GDP) and the active GTP-bound state of Rho (Rho-GTP) both bind to the membrane. They can interconvert through the regulations of GAPs (at rate $J([I_1])$) or GEFs (at rate $B([I_2])$), which may be activated by some external input signals (I_1 and I_2). Rho-GDP can be released from the membrane by binding to a GDI molecule (at rate gdi_R) and revert to its membrane-bound state by releasing GDI (at rate $dgdi\_R$). Rho-GDP and Rho-GTP degrade at rate K_R and K_{R^*} respectively, while the degradation of Rho-GDI was not considered, because GDI binding can stabilize the Rho protein[42]. **(b)** The RhoA-Rac1 regulatory circuit. The GTP-bound states of RhoA or Rac1 can promote GTP loading of its own, and meanwhile activate the GTP hydrolysis of the other. RhoA-GTP is also transcriptionally self-activated. Grb2 inducees the GTP loading of Rac1, while Gab1 induces that of both Rac1 and RhoA. **(c)** The effective (reduced Rac1/RhoA) circuit. In terms of Rac1-GTP and Rho-GTP, their mutual inhibitions form a non-canonical toggle switch with positive auto-

regulations. A solid arrow denotes activation, a solid bar indicates repression, and the wavy line represents regulation by external signals. The solid double line represents the GTP loading or hydrolysis process while the dashed double one represents the binding or unbinding process of GDI molecules. The dotted lines indicate the indirect regulations on GTP loading or hydrolysis process via GEFs or GAPs. **(d)** Typical values of the B and the J functions with respect to the concentrations of the GTPases. The B and J functions represent the GTP loading and hydrolysis rates of both Rac1 and RhoA respectively. Both functions increase with the level of GTP-bound Rac1 and RhoA.

Here, we show in detail how we derived the model from binding and unbinding reactions at molecular level. We first focused on the process of cycling for a typical Rho GTPase (denoted as Rho) (Figure 2.3). The reactions for this process are

$$\text{Rho-DNA} \xrightarrow{k_1} \text{Rho} \quad (1)$$

$$\text{Rho} + \text{GDP} \xrightarrow{k_2} \text{Rho-GDP} \quad (2)$$

$$\text{Rho-GDP} + \text{GDI} \underset{k_4}{\overset{k_3}{\rightleftharpoons}} \text{Rho-GDI} \quad (3)$$

$$\text{Rho-GDP} + \text{GTP} \xrightarrow{k_5} \text{Rho-GTP} + \text{GDP} \quad (4)$$

$$\text{Rho-GDP} + \text{GTP} + \text{GEF} \xrightarrow{k_6} \text{Rho-GTP} + \text{GDP+GEF} \quad (5)$$

$$\text{Rho-GTP} \xrightarrow{k_7} \text{Rho-GDP} + \text{Pi} \quad (6)$$

$$\text{Rho-GTP} + \text{GAP} \xrightarrow{k_8} \text{Rho-GDP} + \text{Pi+GAP} \quad (7)$$

$$\text{Rho-GDP} \xrightarrow{k_9} \text{Null} \quad (8)$$

$$\text{Rho-GTP} \xrightarrow{k_{10}} \text{Null} \quad (9)$$

where Rho-DNA represents the gene encoding this Rho GTPase, and Rho-GTP, Rho-GDP and Rho-GDI stand for Rho GTPase bound to GTP, GDP and GDI respectively.

Reaction (1) is for the production of Rho GTPase. Reaction (2) represents our assumption that Rho protein first binds to GDP instead of GTP when it is translated. Reaction (3) is the reversible binding and releasing of GDI to Rho protein. Reaction (4) is the intrinsic GTP loading process and Reaction (5) is the activated GTP loading by GEFs. Reaction (6) is the intrinsic GTP hydrolysis process and Reaction (7) is the activated GTP hydrolysis by GAPs. At first, we did not consider the signals (I_1, I_2) that can regulate some GEFs or GAPs, thus [GEF] and [GAP] involved in the regulations of the GTP loading and hydrolysis are constant. Reaction (8) and (9) stand for the degradations of Rho-GDP and Rho-GTP, respectively. Since GDI is reported to stabilize Rho GTPases, we do not consider the degradation of Rho-GDI [42]. k_n (n = 1,2 …) is the reaction rate constant for each reaction. The model for these reactions can be given as:

$$\begin{aligned}
\frac{d[R^f]}{dt} &= k_1 \bullet [R^g] - k_2 \bullet [GDP] \bullet [R^f] \\
\frac{d[R^I]}{dt} &= k_3 \bullet [GDI] \bullet [R] - k_4 \bullet [R^I] \\
\frac{d[R]}{dt} &= k_2 \bullet [GDP] \bullet [R^f] + k_4 \bullet [R^I] + k_7 \bullet [R^*] + k_8 \bullet [GAP] \bullet [R^*] \\
&\quad - k_3 \bullet [GDI] \bullet [R] - k_5 \bullet [GTP] \bullet [R] - k_6 \bullet [GTP] \bullet [GEF] \bullet [R] - k_9 \bullet [R] \\
\frac{d[R^*]}{dt} &= k_5 \bullet [GTP] \bullet [R] + k_6 \bullet [GTP] \bullet [GEF] \bullet [R] - k_7 \bullet [R^*] - k_8 \bullet [GAP] \bullet [R^*] - k_{10} \bullet [R^*]
\end{aligned} \quad , (5)$$

where R^f and R^g represent the Rho GTPase without binding to any molecules and the gene encoding Rho GTPase respectively. In order to simplify the equations, we proposed three assumptions here:

1) The levels of GDI, GTP and GDP are assumed to be constant, thus we can reduce some parameters by integrating the number of [GDI], [GTP] and [GDP] into the reaction rate constants (such as $k_3^{'} = k_3 \bullet [GDI]$).

2) The levels of each GEF and GAP are constant except the ones activated by the signals (I_1, I_2) we discussed later. Therefore, we can also integrate them into the reaction rate constants (such as $k_8^{'} = k_8 \bullet [GAP]$).

3) We assume that GDP is present in abundant amount. Rho can bind to GDP to produce inactive GDP-bound form (Rho-GDP) as soon as Rho protein is produced. Thus the $\frac{d[R^f]}{dt}$ equals to 0.

With these assumptions, the model can be converted to:

$$\begin{aligned}
\frac{d[R^I]}{dt} &= k_3^{'} \bullet [R] - k_4 \bullet [R^I] \\
\frac{d[R]}{dt} &= G + k_4 \bullet [R^I] + (k_7 + k_8^{'}) \bullet [R^*] \\
&\quad - k_3^{'} \bullet [R] - (k_5^{'} + k_6^{'}) \bullet [R] - k_9 \bullet [R] \\
\frac{d[R^*]}{dt} &= (k_5^{'} + k_6^{'}) \bullet [R] - (k_7 + k_8^{'}) \bullet [R^*] - k_{10} \bullet [R^*], \quad (6)
\end{aligned}$$

where $G = k_1 \bullet [R^g]$. $k_3^{'}$ and k_4 are binding and releasing rate constant for GDI respectively. k_7 and $k_5^{'}$ are intrinsic GTP hydrolysis and loading rate constants respectively, while $k_8^{'}$ and $k_6^{'}$ are activated GTP hydrolysis and loading rate constants due to GAPs and GEFs respectively.

Now we consider signals (I_1, I_2) in the GBC model, as shown in Figure 2.3. These signals can regulate the activity level of Rho GTPase through GEFs or GAPs.

These signals can be external, such as Grb2 and Gab1 from c-Met pathway or other Rho GTPases like RhoA that can promotes the GTP hydrolysis of Rac1-GTP by activating its relevant GAPs; or internal such as the auto-regulations for Rac1 and RhoA. Therefore, the total GTP hydrolysis and loading rate constants (described by B and J functions respectively as given below) can be divided into three parts: the intrinsic ones, the activated ones by the GAPs and GEFs that are independent of the regulations of signals, and the activated ones by the GAPs and GEFs that depend on the signals. The first two parts, which can be grouped as basal GTP hydrolysis and loading rate constants, are both constant. Yet, the latter activated rate constants (k_A) are related with the level of signals (I) and can be described by Hill function as:

$$H^+([I]) = \frac{[I]^n}{K_d^n + [I]^n} = \frac{k_A}{k_A^{\max}}, \quad (7)$$

where $k_A^{\max}$ is the maximum activated rate constant. K_d is the dissociation constant which represents the threshold at which the activated rate (k_A) is at half maximum value. n is the Hill coefficient determining the steepness of the function. Thus, the activated GTP hydrolysis and loading rate constants by signal I_1 and I_2 can be expressed as $k_A^{GAP_{\max}} \bullet H^+([I_1])$ and $k_A^{GEF_{\max}} \bullet H^+([I_2])$, respectively. Considering these signals into the model above (Equation (6)), the signals-driven model can be given as:

$$\begin{aligned}
\frac{d[R^I]}{dt} &= k_3^{'} \bullet [R] - k_4 \bullet [R^I] \\
\frac{d[R]}{dt} &= G + k_4 \bullet [R^I] + (k_7 + k_8^{"} + k_A^{GAP_{\max}} \bullet H^+([I_1])) \bullet [R^*] \\
&\quad - k_3^{'} \bullet [R] - (k_5^{'} + k_6^{"} + k_A^{GEF_{\max}} \bullet H^+([I_2])) \bullet [R] - k_9 \bullet [R] \\
\frac{d[R^*]}{dt} &= (k_5^{'} + k_6^{"} + k_A^{GEF_{\max}} \bullet H^+([I_2])) \bullet [R] - (k_7 + k_8^{"} + k_A^{GAP_{\max}} \bullet H^+([I_1])) \bullet [R^*] - k_{10} \bullet [R^*]
\end{aligned} \quad , (8)$$

where we use $k_8^{"}$ and $k_6^{"}$ to represent the signal-independent activated rate constants for GTP hydrolysis and loading, respectively. To simplify Equation (8), we defined two functions: one is B function standing for total GTP loading rate constant and the other is J function standing for total GTP hydrolysis rate constant. Both of them depend on the signals.

$$\begin{aligned}
B([I_2]) &= k_5^{'} + k_6^{"} + k_A^{GEF_{\max}} \bullet H^+([I_2]) \\
J([I_1]) &= k_7 + k_8^{"} + k_A^{GAP_{\max}} \bullet H^+([I_1])
\end{aligned} \quad , (9)$$

Integrating Equation (8) and (9), we can get a generic deterministic model (Equation (10)) to describe the transition among GDI state, GDP state and GTP state of a typical Rho GTPase (Figure 2.3).

$$\begin{aligned}
\frac{d[R^I]}{dt} &= k_3^{'} \bullet [R] - k_4 \bullet [R^I] \\
\frac{d[R]}{dt} &= G + k_4 \bullet [R^I] + J([I_1]) \bullet [R^*] - k_3^{'} \bullet [R] - B([I_2]) \bullet [R] - k_9 \bullet [R] \\
\frac{d[R^*]}{dt} &= B([I_2]) \bullet [R] - J([I_1]) \bullet [R^*] - k_{10} \bullet [R^*]
\end{aligned} \quad , (10)$$

Mathematical Model of Rac1/RhoA regulatory circuit

We utilized the approach to investigate the dynamics of the core Rac1/RhoA regulatory circuit, as shown in Figure 2.3. The deterministic dynamics of the circuit could be modeled by a set of six chemical rate equations presented as below:

$$
\begin{aligned}
&\frac{d[R_c^I]}{dt} = gdi\_R_c \bullet [R_c] - dgdi\_R_c \bullet [R_c^I] \\
&\frac{d[R_c]}{dt} = g_{R_c} + J_{R_c}([R_h^*]) \bullet [R_c^*] + dgdi\_R_c \bullet [R_c^I] \\
&\qquad - gdi\_R_c \bullet [R_c] - B_{R_c}([R_c^*]) \bullet [R_c] - I_{R_c} \bullet [R_c] - K_{R_c} \bullet [R_c] \\
&\frac{d[R_c^*]}{dt} = B_{R_c}([R_c^*]) \bullet [R_c] + I_{R_c} \bullet [R_c] - J_{R_c}([R_h^*]) \bullet [R_c^*] - K_{R_c^*} \bullet [R_c^*] \\
\\
&\frac{d[R_h^I]}{dt} = gdi\_R_h \bullet [R_h] - dgdi\_R_h \bullet [R_h^I] \\
&\frac{d[R_h]}{dt} = (g_{R_h} + g_{R_hA} \bullet H^+([R_h^*])) + J_{R_h}([R_c^*]) \bullet [R_h^*] + dgdi\_R_h \bullet [R_h^I] \\
&\qquad - gdi\_R_h \bullet [R_h] - B_{R_h}([R_h^*]) \bullet [R_h] - I_{R_h} \bullet [R_h] - K_{R_h} \bullet [R_h] \\
&\frac{d[R_h^*]}{dt} = B_{R_h}([R_h^*]) \bullet [R_h] + I_{R_h} \bullet [R_h] - J_{R_h}([R_c^*]) \bullet [R_h^*] - K_{R_h^*} \bullet [R_h^*]
\end{aligned}
\quad , (11)
$$

where

$$
\begin{aligned}
&B_{R_c}([R_c^*]) = gtp\_R_c i + gtp\_R_c B + gtp\_R_c A \bullet H^+([R_c^*]) \\
&J_{R_c}([R_h^*]) = dgtp\_R_c i + dgtp\_R_c B + dgtp\_R_c A \bullet H^+([R_h^*]) \\
&I_{R_c} = (gtp\_R_c I_1 \bullet H^+([I_1]) + gtp\_R_c I_2 \bullet H^+([I_2])) \\
\\
&B_{R_h}([R_h^*]) = gtp\_R_h i + gtp\_R_h B + gtp\_R_h A \bullet H^+([R_h^*]) \\
&J_{R_h}([R_c^*]) = dgtp\_R_h i + dgtp\_R_h B + dgtp\_R_h A \bullet H^+([R_c^*]) \\
&I_{R_h} = gtp\_R_h I_2 \bullet H^+([I_2])
\end{aligned}
\quad , (12)
$$

Here, R_c^I, R_c, R_c^* stand for Rac1 in GDI-bound state, GDP-bound state, and GTP-bound state, respectively, R_h^I, R_h, R_h^* stand for the different states for RhoA. B

function is the total GTP loading rate constant, which contains the intrinsic GTP loading rate constant ($gtp\_R_c i$ and $gtp\_R_h i$), the activated GTP loading rate constant ($gtp\_R_c B$ and $gtp\_R_h B$) by GEFs not involved in the auto-regulations and the activated GTP loading rate constant ($gtp\_R_c A$ and $gtp\_R_h A$) resulting from auto-regulations. J function is the total GTP hydrolysis rate constant, which also contains the intrinsic GTP hydrolysis rate constant ($dgtp\_R_c i$ and $dgtp\_R_h i$), the activated GTP hydrolysis rate constant ($dgtp\_R_c B$ and $dgtp\_R_h B$) by GAPs not involved in the mutually regulations and the activated GTP hydrolysis rate constant ($dgtp\_R_c A$ and $dgtp\_R_h A$) resulting from the mutual regulations. gdi_R_c , $dgdi\_R_c$, gdi_R_h , and $dgdi\_R_h$ are the binding and unbinding rate constants for GDI. There are two external signals here: Grb2 (I_1) and Gab1 (I_2). I_{R_c} is the activated GTP loading rate constant for Rac1 by both Grb2 ($gtp\_R_c I_1$) and Gab1 ($gtp\_R_c I_2$) signals, and I_{R_h} is the activated GTP loading rate constant for RhoA by Gab1 ($gtp\_R_h I_2$) signal.

As the GTP-bound states of RhoA and Rac1 are the only active regulators of cell migration, we focus on the concentration of Rac1-GTP and RhoA-GTP and reduce the equations to two coupled equations for R_c^* and R_h^*. Assuming the total expression of Rac1 and RhoA reach equilibrium with their degradation, therefore

$$\frac{d[R_{ctot}]}{dt} = \frac{d[R_c^I]}{dt} + \frac{d[R_c]}{dt} + \frac{d[R_c^*]}{dt} = 0$$
$$\frac{d[R_{htot}]}{dt} = \frac{d[R_h^I]}{dt} + \frac{d[R_h]}{dt} + \frac{d[R_h^*]}{dt} = 0 \quad , (13)$$

Then we can get

$$g_{R_c} - K_{R_c} \bullet [R_c] - K_{R_c^*} \bullet [R_c^*] = 0$$

$$(g_{R_h} + g_{R_hA} \bullet H^+([R_h^*])) - K_{R_h} \bullet [R_h] - K_{R_h^*} \bullet [R_h^*] = 0 \quad , (14)$$

If K_{R_c} (K_{R_h}) is same as $K_{R_c^*}$ ($K_{R_h^*}$), $[R_c]$ and $[R_h]$ can be expressed as a function of $[R_c^*]$ and $[R_h^*]$:

$$[R_c] = \frac{g_{R_c}}{K_{R_c^*}} - [R_c^*]$$

$$[R_h] = \frac{(g_{R_h} + g_{R_hA} \bullet H^+([R_h^*]))}{K_{R_h^*}} - [R_h^*] \quad , (15)$$

By substituting these expressions back into Equation (11), we simplified the original model to four equations as below:

$$\frac{d[R_c^I]}{dt} = gdi\_R_c \bullet [R_c] - dgdi\_R_c \bullet [R_c^I]$$
$$\frac{d[R_c^*]}{dt} = \frac{g_{R_c}}{K_{R_c^*}} \bullet (B([R_c^*]) + I_{R_c}) - (B([R_c^*]) + J([R_h^*]) + I_{R_c} + K_{R_c^*}) \bullet [R_c^*]$$
$$\frac{d[R_h^I]}{dt} = gdi\_R_h \bullet [R_h] - dgdi\_R_h \bullet [R_h^I]$$
$$\frac{d[R_h^*]}{dt} = \frac{(g_{R_h} + g_{R_hA} \bullet H^+([R_h^*]))}{K_{R_h^*}} \bullet (B([R_h^*]) + I_{R_h}) - (B([R_h^*]) + J([R_c^*]) + I_{R_h} + K_{R_h^*}) \bullet [R_h^*] \quad , (16)$$

Since only active Rac1 and RhoA cause downstream reactions, we can focus on the two-coupled equations with variable $[R_c^*]$ and $[R_h^*]$ as an effective model to approximate the detail model.

$$\frac{d[R_c^*]}{dt} = \frac{g_{R_c}}{K_{R_c^*}} \bullet (B([R_c^*]) + I_{R_c}) - (B([R_c^*]) + J([R_h^*]) + I_{R_c} + K_{R_c^*}) \bullet [R_c^*]$$

$$\frac{d[R_h^*]}{dt} = \frac{(g_{R_h} + g_{R_hA} \bullet H^+([R_h^*]))}{K_{R_h^*}} \bullet (B([R_h^*]) + I_{R_h}) - (B([R_h^*]) + J([R_c^*]) + I_{R_h} + K_{R_h^*}) \bullet [R_h^*] \quad , (17)$$

where the first part such as $\frac{g_{R_c}}{K_{R_c^*}} \bullet (B([R_c^*]) + I_{R_c})$ is the effective production rate, while the second part such as $(B([R_c^*]) + J([R_h^*]) + I_{R_c} + K_{R_c^*}) \bullet [R_c^*]$ is the effective degradation part. Notice that the self-activation on GTP loading process can activate both the effective production rate and the effective degradation rate. These simplified equations describe the interaction between active Rac1 and RhoA more clearly, and can be used to determine the stability of Rac1/RhoA regulatory circuit. The effective model was used for stability and bifurcation analysis while the detailed model was used for the dynamic simulations.

Mathematical Model of the Coupled Circuit

The EMT regulatory circuit regulates the Rac1/RhoA circuit via miR-34 and miR-200 (Figure 2.2b). Here, we use an effective (reduced) model (Figure 2.8a) to represent the association between these two circuits – miR-34/SNAIL/miR-200/ZEB (the core circuit for EMT/MET) and RhoA/Rac1 (the core circuit for AMT/MAT) (Figure 2.2b). As shown in Figure 2.8a, the effect of EMT regulatory circuit is treated as two external signals to the Rac1/RhoA circuit, where (μ1) represents the inhibition on RhoA by either of the two microRNAs (miR-34 or miR-200) while (μ2) represents a similar inhibition on Rac1.

We have previously developed the theoretical framework for microRNA-based regulatory circuit like miR-34/SNAIL and miR-200/ZEB. Here, we use both these frameworks to study the dynamics of the coupled effective circuit. The deterministic dynamics of the circuit can be modeled by two rate equations as below:

$$\frac{d[R_c^*]}{dt} = \frac{g_{R_c}}{K_{R_c^*}} P_{R_c}([\mu_2])(B_{R_c}([R_c^*]) + I_{R_c}) - (B_{R_c}([R_c^*]) + J_{R_c^*}([R_h^*]) + I_{R_c} + K_{R_c^*})[R_c^*]$$
$$\frac{d[R_h^*]}{dt} = \frac{(g_{R_h} + g_{R_hA} H^+_{R_h^*}([R_h^*]))}{K_{R_h^*}} P_{R_h}([\mu_1])(B_{R_h}([R_h^*]) + I_{R_h}) - (B_{R_h}([R_h^*]) + J_{R_h^*}([R_c^*]) + I_{R_h} + K_{R_h^*})[R_h^*] \quad , \quad (18)$$

where R_C^* represents the active Rac1 (Rac1-GTP), and R_h^* represents the active RhoA (RhoA-GTP). g_{R_c} is the production rate for Rac1, g_{R_h} and g_{R_hA} are basal and excitatory production rates for RhoA respectively. $K_{R_c^*}$ and $K_{R_h^*}$ are the corresponding degradation rates for Rac1 and RhoA. I_{R_c} and I_{R_h} represent two external signals that drive the GTP hydrolysis/loading process of the Rac1/RhoA circuit, such as Grb2 and Gab1 in c-MET signaling, whose effects have been analyzed in our previous work[38]. The Hill function $H^+_{R_h}([R_h^*])$ represents the transcriptional self-activation of RhoA. B and J functions are defined to represent the total GTP loading and hydrolysis rates including the intrinsic ones for RhoA and Rac1 and also the activated ones by GEFs or GAPs (See above)[38]. The inhibition effect of microRNAs is described by the function $P([\mu])$ whose value ranges from 0 to 1 – 0 denotes a strong silencing effect, and 1 denotes no effect[11].

Parameters estimation

The values of most of the parameters considered in our model are not known exactly, but we hereby explain how we estimated the parameter values. The detail parameters are summarized in Table S1.2.

The degradation rates are selected due to the half-lives of each molecule from experiments. The half-life for Rac1 is about 2 hours[59], while the half-life of RhoA is about 5 hours[60]. Thus, we set their degradation rate to be 0.1 h^{-1}.

The total amount of Rac1 and RhoA protein are respectively 82-123 ng/10^6 cells and 38-75 ng/10^6 cells. These values are averaged from two epithelial cell lines, MDCK and ECV[61]. MDCK and ECV have similar amount of Rac1 and RhoA, where MDCK cells have about 56±14 ng/10^6 cells RhoA and 124±27 ng/10^6 cells Rac1 while ECV304 cells have about 50±15 ng/10^6 cells RhoA and 82±14 ng/10^6 cells Rac1. We used the average values of them here to ensure validity of our parameters. Since the molecular weight of these small G-proteins is about 22 kDa (3.65×10^{-11}ng), the molecular numbers of Rac1 and RhoA per cell are 2.25×10^6 -3.37×10^6 molecules and 1.04×10^6-2.05×10^6 molecules, respectively [61]. Thus, we set the basal transcriptional rates for Rac1 (g_{R_c}) and RhoA (g_{R_h}) to 3.40×10^5 molecules/hour and 1.60×10^5 molecules/hour. Also, due to the transcriptional self-activation by RhoA, we assume that for the cells in amoeboid state, they can have high expression of RhoA similar to that of Rac1 in cells with mesenchymal morphology. Thus, the activated transcriptional rate (g_{R_hA}) for RhoA is also set to 3.40×10^5 molecules/hour.

GTPases have a slow intrinsic GTP hydrolysis and loading ability. Experiments show that the intrinsic GTP hydrolysis rates for Rac1 and RhoA are about 6.6 h^{-1} and 1.32 h^{-1}, respectively[62,63]. GAPs can accelerate the hydrolysis rates by 5 to 4000 folds[63,64]. Thus we set $dgtp\_R_cB$ to 110 h^{-1} and $dgtp\_R_cA$ to 200 h^{-1} for Rac1, and set $dgtp\_R_hB$ to 311 h^{-1} and $dgtp\_R_hA$ to 88 h^{-1} for RhoA. For GTP loading process, the intrinsic rates are both about 0.54 h^{-1} for active Rac1 and RhoA[65]. Experiments show that GEFs can accelerate the rates by 3.5 to 1000 folds[65–67]. Similarly, we set gtp_R_cB to 20 h^{-1} and gtp_R_cA to 530 h^{-1} for Rac1, and set gtp_R_hB to 110 h^{-1} and gtp_R_hA to 196 h^{-1} for RhoA. Also, we take 2000 h^{-1} for both GDI binding (gdi_R_c and gdi_R_h) and releasing ($dgdi\_R_c$ and $dgdi\_R_h$) rate constants because Rac1 was observed to stay on the membrane for 2s[68].

In addition, since regulations such as the auto-regulations and the mutual inhibitions between Rac1-GTP and RhoA-GTP are indirect, we considered Hill function coefficients to be 4. Also, for Grb2 and Gab1 signals, we integrated them in the model using hill functions. Since their concentration is suggested to be about 1μM[69], namely about 1×10^6 molecules by considering the typical diameter of a eukaryotic cell to be 10μm, we set their corresponding hill function thresholds ($Tsd\_gtp\_R_cI_1$, $Tsd\_gtp\_R_hI_2$ and $Tsd\_gtp\_R_cI_2$) to 5×10^5 molecules. Since we integrated these signals into our model by hill functions, it is worthy to note that it is the ratio between the concentration of Gab1 and Grb2 to their thresholds instead of their absolute concentration values that actually affects the effects of these signals. Thus, the absolute values we used for Gab1 and Grb2 signals can always be adopted by modifying the threshold values.

miR-34a could either suppress the transcription of RhoA by targeting its transcriptional complexes or directly Rac1[53,57]. Also, the bioinformatics tool, TargetScan[70–73], indicates the possible inhibition of miR-200 on Rac1 and RhoA, which is consistent with several experimental results[54–56]. To simplify these connections between miR-34/miR-200 and Rac1/RhoA, we here consider both of them directly inhibit the expression of Rac1 and RhoA, and group their inhibitions on RhoA to be ' μ_1 ' while the ones on Rac1 to be ' μ_2 '. The effective binding sites for μ_1 and μ_2 on both RhoA and Rac1 is set to 2, and we use the same parameters for the P functions as we used before[40]. Therefore, the protein expressions of target proteins of these microRNAs come down to 30% of the control case (no inhibition by microRNAs). The thresholds for P function is set to be 10, 000 molecules, but it should be noted that it is the ratio between the concentration of microRNAs to their threshold levels instead of their absolute concentration values that actually governs the effects of these signals.

2.3. The three-way switch operation of Rac1/RhoA GTPase-based circuit controlling amoeboid-hybrid-mesenchymal transition

The Rac1/RhoA Circuit as a Three-way Switch

We started by analyzing the circuit dynamics in the absence of external signals (Grb2 = 0 and Gab1 = 0). As illustrated by a typical phase-plane diagram in Figure 2.4a, Rac1/RhoA regulatory circuit can act as a three-way switch among the following three states: (high active RhoA/low active Rac1), (low active RhoA/high active Rac1), and (both high active RhoA and Rac1), which we denote as (1, 0), (0+, 1) and (1, 1)

respectively. According to the experimental observations[35,44–46,74], we associate the states (1, 0) and (0+, 1) with the amoeboid (A) and mesenchymal (M) phenotypes. Again, we used "0+" to denote some minimal level of active RhoA present in the rear end of mesenchymal cells and required for their individual migration[75]. Furthermore, we proposed to associate the (1, 1) state with the amoeboid-mesenchymal (A/M) hybrid phenotype that has been suggested in some recent experiments both for cancer cells and normal cells during early embryonic development[31,35,37,76–80]. Although these experiments lack quantitative measurement of the activity of Rac1 and RhoA, the mixed morphologies of these cells share the traits of both amoeboid and mesenchymal phenotypes. These properties may be indicative of relative high levels of both active RhoA and Rac1. Yet, it is clear that a direct measurement of these proteins is necessary to establish a direct association between these phenotypes and the model predicted hybrid phenotype.

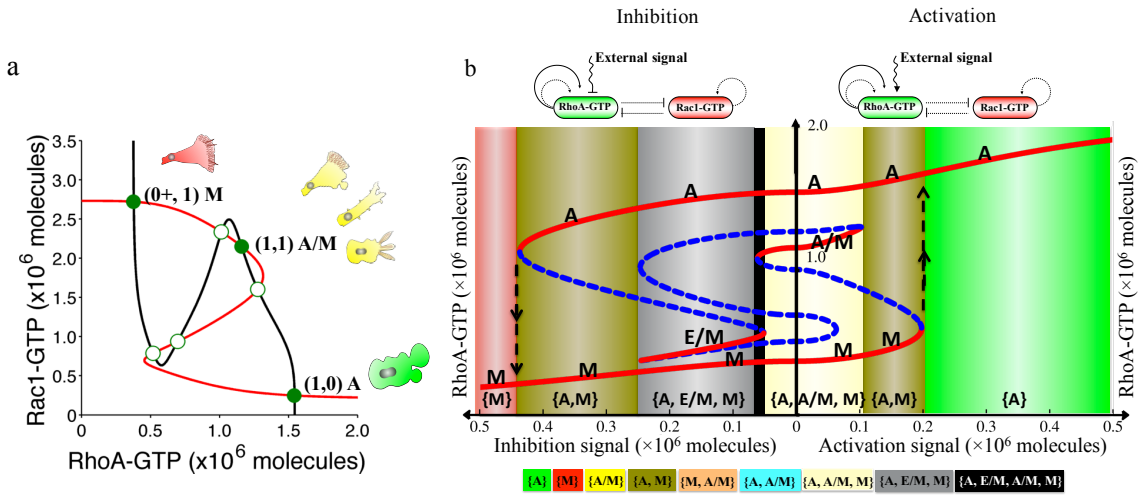


Figure 2.4 Rac1/RhoA circuit as a three-way switch.

(a) Dynamical system characteristics of the Rac1/RhoA regulatory circuit. The plot shows the nullclines and possible steady states corresponding to Equation (17). Without any external signal (Grb2 = 0, Gab1 = 0), the circuit can be tristable. Red nullcline is for $d[R_c^*]/dt = 0$ and black nullcline is for $d[R_h^*]/dt = 0$. Green solid circles denote the stable fixed points, and green hollow circles denote the unstable fixed points. Each stable point can be associated with a cell phenotype, depicted as a cartoon beside them. **(b)** Bifurcation of RhoA-GTP protein levels in response to an external signal regulating the GTP loading rates of RhoA. The signal can either increase (activation) or decrease (inhibition) the GTP loading rate. The response to activation/inhibition is shown on the right/left side of the bifurcation respectively. The red solid lines indicate stable states and the blue dashed lines indicate unstable states. The bifurcation illustrates the possible coexistence (for some range of the signal) of four states: (*i*) the (1, 0) state with high RhoA-GTP and low Rac1-GTP, which corresponds to A phenotype; (*ii*) the (0+, 1), which corresponds to M phenotype; (*iii*) the (1, 1), which correspond to A/M phenotype; (*iv*) the (0, 0) state, which corresponds to E/M phenotype. The corresponding bifurcation of Rac1-GTP protein levels is showed in Figure S1.2. Co-existence of different phenotypes is associated with a multistable phase, highlighted by different background colors (see legend at the bottom). Starting from the (0+, 1) state (M phenotype, at bottom left part of the red curve), the system stays in the M phenotype when the inhibition signal is reduced; further switching the inhibition signal to an increasing activation signal leads the system to undergo a transition to the (1, 0) state (A phenotype, at top right part of the red curve). The transition is indicated by the dashed upward arrow at the boundary of the phase {A, M} and {A}. Similarly, increasing the inhibition signal can induce the transition from the (1, 0) state (A phenotype) back to the (0+, 1) state (M phenotype), as indicated by the downward arrow at the boundary of the phase {M} and {A, M}. Besides, cells may switch to the A/M or E/M phenotype due to noises in gene expression.

The Switch Response to External Activation and Inhibition Signals

Next, we analyzed the response of the Rac1/RhoA circuit to an external input signal that drives the circuit through either Rac1 or RhoA. We modeled that the signal directly increases or decreases the basal GTP loading rates. In Figure 2.4a, we show the

response to the external signal that affects the RhoA loading rate by the effective model. When the signal activates the basal GTP loading rate for RhoA, it gives rise to the coexistence of the diverse phenotypes A, M, A/M, all of which correspond to solitary movement. However, when the signal inhibits the loading rate, it also gives rise to collective cell migration of the E/M hybrid phenotype. As is also evident from Figure 2.4b, a high activation of the GTP loading rate leads the cell to a monostable phase {A} in which only amoeboid phenotype (A) exists, whereas low levels of GTP loading rate correspond to the monostable phase {M} in which only the mesenchymal phenotype (M) exists. Our model is consistent with the experiments showing that cells with constitutively active RhoA are associated with amoeboid (A) or blebby amoeboid (BA) phenotype[81]; while cells with dominant negative RhoA usually exhibit a mesenchymal (M) phenotype[32]. It also predicted that, when the external signal acts only on RhoA, the induced transitions between A and M phenotypes are much easier than the transitions from A or M phenotype to A/M or E/M phenotype. This may explain why these hybrid phenotypes, A/M and E/M, were rarely observed during AMT in most of the experiments[36].

The Switch Response to Input Signals from Grb2 and Gab1

As was mentioned earlier, Grb2 activates only Rac1, and Gab1 activates both Rac1 and RhoA. To understand the circuit response to these regulations, we first investigated the response of the circuit dynamics to either Grb2 or Gab1 (in terms of the corresponding bifurcation diagram) when they act individually. We found that when Grb2 level is increased, the cells adopt a mesenchymal (M) phenotype; whereas Gab1

induces the cell to adopt an amoeboid (A) phenotype (Figure S1.3). However, further high Gab1 signal can induce the cell to adopt the amoeboid/mesenchymal (A/M) phenotype since Gab1 can activate both Rac1 and RhoA (Figure S1.3).

Next, to understand the combined effect of Grb2 and Gab1, we constructed the two-parameter bifurcation diagram, as shown in (Figure 2.5). Each phase corresponds to a particular situation in which one or several different phenotypes can coexist. More specifically, the possible phases are: 1. Phases with only one phenotype – {A}, {M} and {A/M}. 2. Phases in which two phenotypes can coexist – {A, A/M}, {M, A/M} and {A, M}. 3. A phase in which all three phenotypes coexistence – {A, M, A/M}. The detailed information of the various phases indicates the plasticity of cell migration as driven by different combinations of Grb2 and Gab1 signals. Depending on how Grb2 and Gab1 increase temporally, the cells follow different trajectories in this phase diagram and thus go through different phenotypic transitions as is illustrated in Figure 2.5. Understanding the circuit response to input signal from Grb2 and Gab1 provides crucial clues regarding the pleiotropic effects of the c-MET pathway in promoting either the amoeboid or mesenchymal mode of migration and also transitions between them (AMT and MAT).

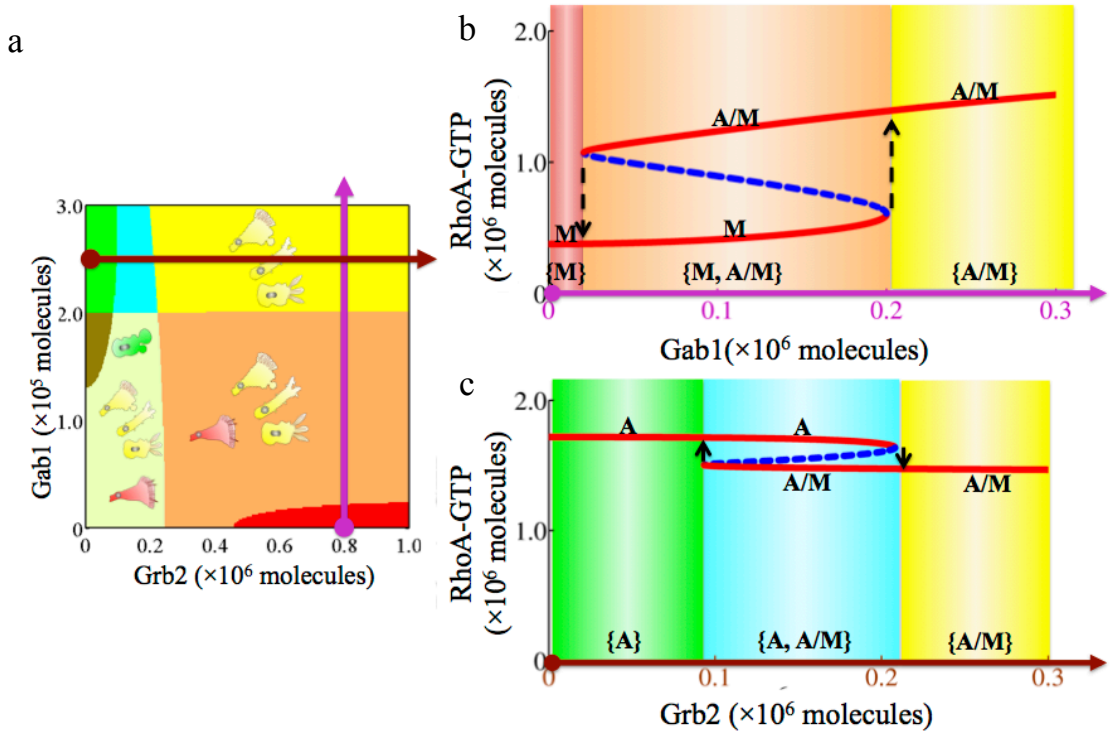


Figure 2.5 The circuit response to input signals from Grb2 and Gab1.

(a) Two-parameter bifurcation phase diagram. Two input signals Grb2 (x-axis) and Gab1 (y-axis) are selected as the two bifurcation parameters. As explained in the text, each phase corresponds to a different combination of coexisting phenotypes (Phase plane diagrams for each phase are showed in Figure S1.4). For example, the orange area is the phase {M, A/M}, which means that cells in this phase can belong to either the M or the A/M phenotype. The colors used for different phases are explained by the legend in Figure 2.4. **(b)** One-parameter bifurcation diagram for the circuit driven by Gab1 when the Grb2 level is fixed to be 8×10^5 molecules. The corresponding trajectory in the two-parameter bifurcation phase diagram (Panel a) is shown by a purple line. The transitions between the M and A/M phenotypes are illustrated by the dashed upward and downward arrows. **(c)** One-parameter bifurcation diagram for the circuit driven by Grb2 when the Gab1 level is fixed to be 2.5×10^5 molecules. The corresponding trajectory in the two-parameter bifurcation phase diagram (Panel a) is shown as a brown line. The transitions between the A and A/M phenotypes are illustrated by the upward and downward arrows.

Phenotypic Transitions Driven by the c-MET pathway

To better understand the phenotypic transitions, we investigated the response dynamics of the Rac1/RhoA circuit when the input signals, Gab1 and Grb2, change in time. To recapitulate the possible response of the circuit to HGF/SF treatment, we chose time dependent functions for the levels of the Gab1 and Grb2 to mimic the cells' response to HGF/SF treatment (Figure 2.6). The HGF/SF signal leads to increase c-MET phosphorylation, which in turn regulates Gab1 and Grb2 together with Met-Induced Mitochondrial Protein (Mimp) in a form of two coupled feed-forward loops (FFLs) as is shown in Figure 2.2[82,83]. Hence, the levels of Gab1 and Grb2 were modeled as two pulse signals with a time delay as shown in Figure 2.6[84].

The form of the pulses is similar to that of a typical response of a FFL. Time delay is incorporated to reflect the effect of the feed-forward like coupling of Grb2 and Gab1 to c-MET (Gab1 responses ahead of Grb2). The simulated treatment starts with cell in the amoeboid (A) phenotype at the left bottom corner (shown as blue star) in the phase diagram shown in Figure 2.6. The cell stays in this phenotype when Gab1 is increased (Figure 2.6) but makes a transition into the hybrid (A/M) phenotype after Gab1 decreases and Grb2 increases, thus the activity of Rac1 increases while that of RhoA remains almost unchanged. Finally, after Grb2 also decreases, the cell goes through another transition from the hybrid (A/M) phenotype into the mesenchymal (M) one. The results illustrate how c-MET pathway can regulate the cells to switch between different migrating phenotypes. In Figure 2.6, we demonstrated the result for the same simulation but for a cell with different circuit parameters (a cell in which Gab1 activation of Rac1

GTP loading is stronger than its activation of the RhoA loading). In this case, the increase of Gab1 signal induces the cell to transit into the hybrid A/M phenotype instead of maintaining it in A phenotype. The different signal-response behaviors of these two simulations support that different cell lines (cells with different parameters) may respond differently to the regulatory signals Gab1 and Grb2.

The aforementioned results are consistent with experimental results and help to explain several experimental observations. In particular, c-MET pathway is reported to induce both the mesenchymal phenotype for several non-cancer and cancer cells[50,51,85] and the amoeboid phenotype for some breast cancer cells[52]. Other experiments show that Grb2 is essential for TGF-beta to induce a mesenchymal phenotype for some cancer cells[86] and Gab1 can stimulate AMT by forming dorsal ruffles through its adaptor Nck[87]. These observations have been well captured by our simulations showing that mesenchymal phenotype can be induced either by Grb2 in one cell line or by Gab1 in another cell line with different parameters (Figure 2.6).

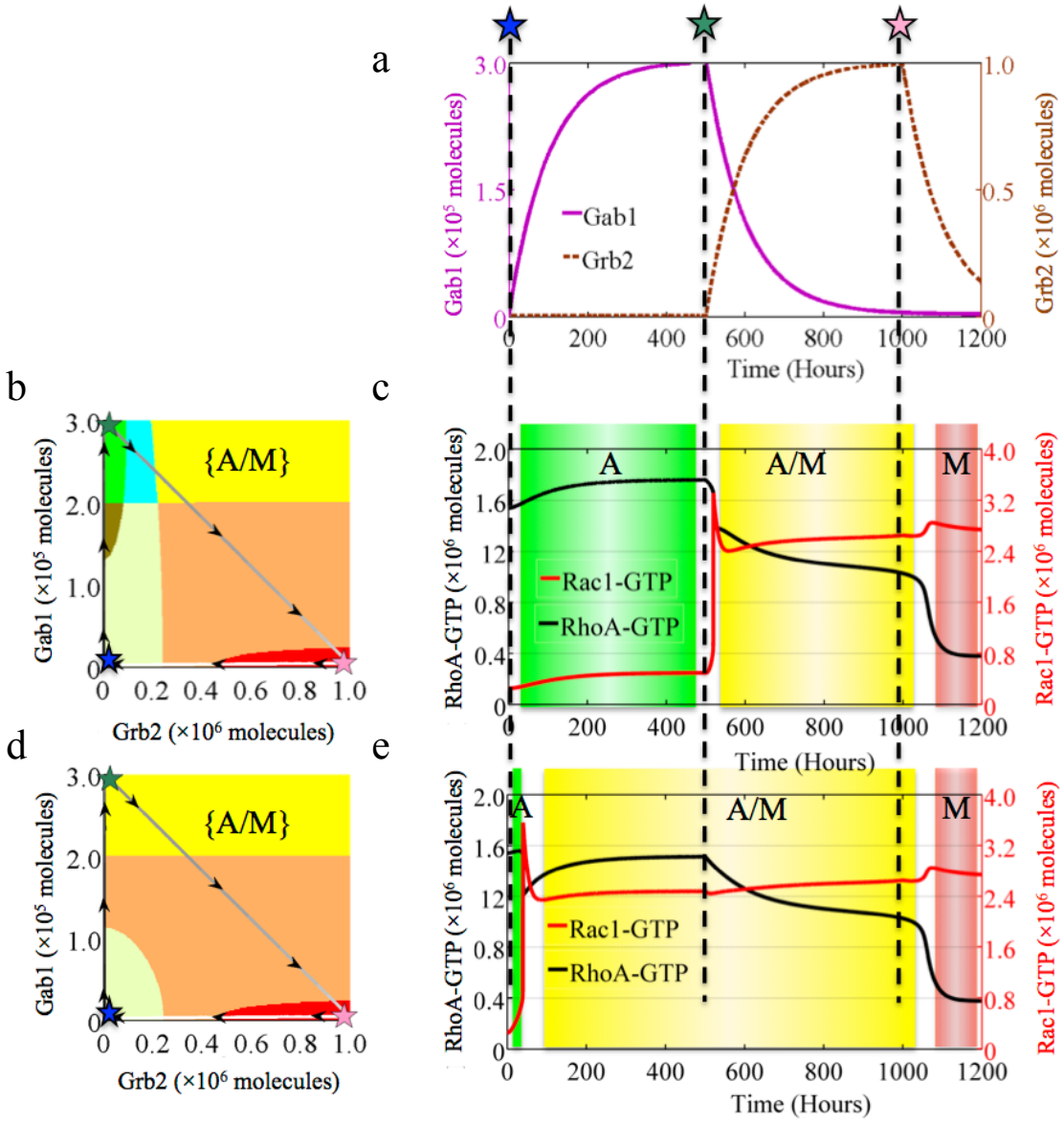


Figure 2.6 Temporal dynamics of the cells to HGF/SF treatment.

(a) The time dependent Gab1 and the Grb2 signals that imitate the effect of HGF/SF treatment, as explained in the text. **(b)** Trajectory of the cell response projected on the phase diagram. The results are for a cell in which Gab1 activation of RhoA GTP loading is stronger than its activation of the Rac1 GTP loading (The parameters are $gtp\_R_hI_2 = 240h^{-1}$ and $gtp\_R_cI_2 = 90h^{-1}$ respectively). The solid line is the trajectory, and both the arrows on the line and the color gradient of the line (from black to white) indicate the time evolution. The blue star marks the initial condition, the green star marks the peak of Gab1 expression and the pink star marks the peak of Grb2 expression. **(c)**

Time dynamics of the expression levels of Rac1-GTP (Red) and RhoA-GTP (Black) in response to the signals, Gab1 and Grb2. The different phenotypes during the transition are highlighted by different background colors, where green, yellow and red areas stand for A, A/M and M phenotype respectively. **(d)** Similar to (b) but for a cell in which Gab1 activation of Rac1 loading is stronger than its activation of the RhoA loading (The parameters are $gtp\_R_cI_2 = 1000h^{-1}$ and $gtp\_R_hI_2 = 240h^{-1}$ respectively). **(e)** Similar to (c) but for the case shown in (d).

Phenotype Distribution

Cells belonging to the same cell line often display non-heritable phenotypic variability[88,89]. Such variability can originate, for example, from local differences in the microenvironment (such as ECM rigidity) leading to differences in the circuit parameters of the individual cells. We have shown earlier that cells with different circuit parameters can respond in a different way to the input signals. Hence, we expect to see a distribution of phenotype for given input signals. As a first step to assess the expected nature of the population level distribution, we extended our simulations to a population of 5,000 cells, each with different circuit parameters. More specifically, the cells that compose the populations have ±5% variations from the original parameters.

In Figure 2.7, we showed the percentages of cells that can be in one of the three different possible phenotypes, (A, A/M and M), for different levels of the input signals. We found that for high Grb2 or high Gab1 signal, a significant percentage of cells adopt the A/M phenotype (Figure 2.7). This result, which is obtained due to a weak robustness (high sensitivity to the circuit parameters), indicates that cells under high Grb2 or Gab1 signal are still sensitive to the conditions of their microenvironment. However, both high

Grb2 and Gab1 signals can result in more cells being in the hybrid A/M phenotype (Figure 2.7).

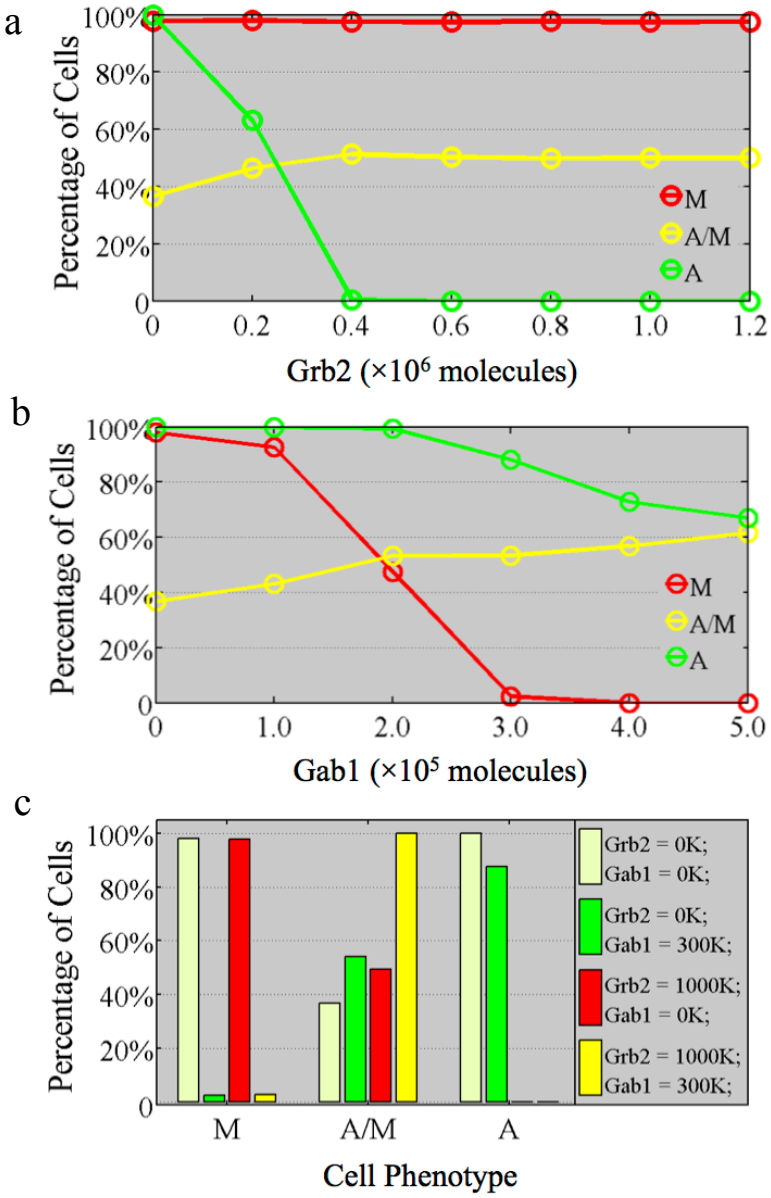


Figure 2.7 Phenotype distribution of a population of cells.

(a) Phenotype distribution of a population of cells driven by Grb2 signal (Gab1 = 0). The cell parameters are randomly distributed over ±5% relative to the original parameters.

The y-axis denotes the percentage of cells corresponding to a specific phenotype. The color of each line represents a phenotype. Under the stimulation of different level of Grb2, a population of cells has different phenotype distribution (percentage of cells showed in y-axis). For each point, the original phase of the cells depends on the level of Grb2 signal. Note that due to the co-existence of different phenotypes in one phase (e.g. in both the phases {A, M, A/M} and {A, M}; A and M phenotypes are present), the sum of the total percentages for one particular signal level can be more than 100%. For example, the initial phase with neither Gab1 nor Grb2 signals is {A, A/M, M}. About 100% cells can be A or M phenotype, and about 36% cells can be A/M. Some of these cells were in some multistable regime comprised of one or more of the phenotypes of A, M and A/M. **(b)** Phenotype distribution driven by Gab1 signal (Grb2 = 0). **(c)** Phenotype distribution under Grb2 and Gab1 regulations. The color of each bar, corresponding to the color definition above represents the initial cell phase. When both Grb2 and Gab1 signals are high, the cells are highly likely to be maintained in hybrid A/M phenotype.

2.4. Modeling the Transitions between Collective and Solitary Migration Phenotypes in Cancer Metastasis

The additional stable steady state corresponding to collective migration phenotype

In the first study, we found that the standalone Rac1/RhoA circuit (i.e. without any input from miR-34 and miR-200) can act as a three-way switch enabling these three states: (a) (high active RhoA/low active Rac1), denoted by (1, 0), (b) (low active RhoA/high active Rac1) denoted by (0+, 1) and (c) (both high active RhoA and Rac1) denoted by (1, 1). Based on experimental observations[31,35–37], these states can be associated with the amoeboid (A), mesenchymal (M) and hybrid amoeboid/mesenchymal (A/M) phenotypes respectively.

Now, considering the connections between Rac1/RhoA circuit and EMT regulatory circuit, i.e. in presence of low levels of microRNAs (miR-34 and miR-200),

the system can display another stable state that has relatively low levels of both active RhoA and Rac1, as denoted by (0, 0) (Figure 2.8b). At certain levels of microRNAs, the system can also behave as a four-way switch such that all these steady states—(0, 0), (1, 0), (1, 1), and (0+, 1)– co-exist (Figure 2.8c). We hereby propose to associate this new (0, 0) state with the hybrid epithelial/mesenchymal (E/M) phenotype. This association is corroborated by experimental observations of the presence of balanced, moderate levels of active Rac1 and RhoA in cells of the hybrid E/M phenotype not only during cancer metastasis (type III EMT), but also during wound healing (type II EMT), and embryonic development (type I EMT)[90–92].

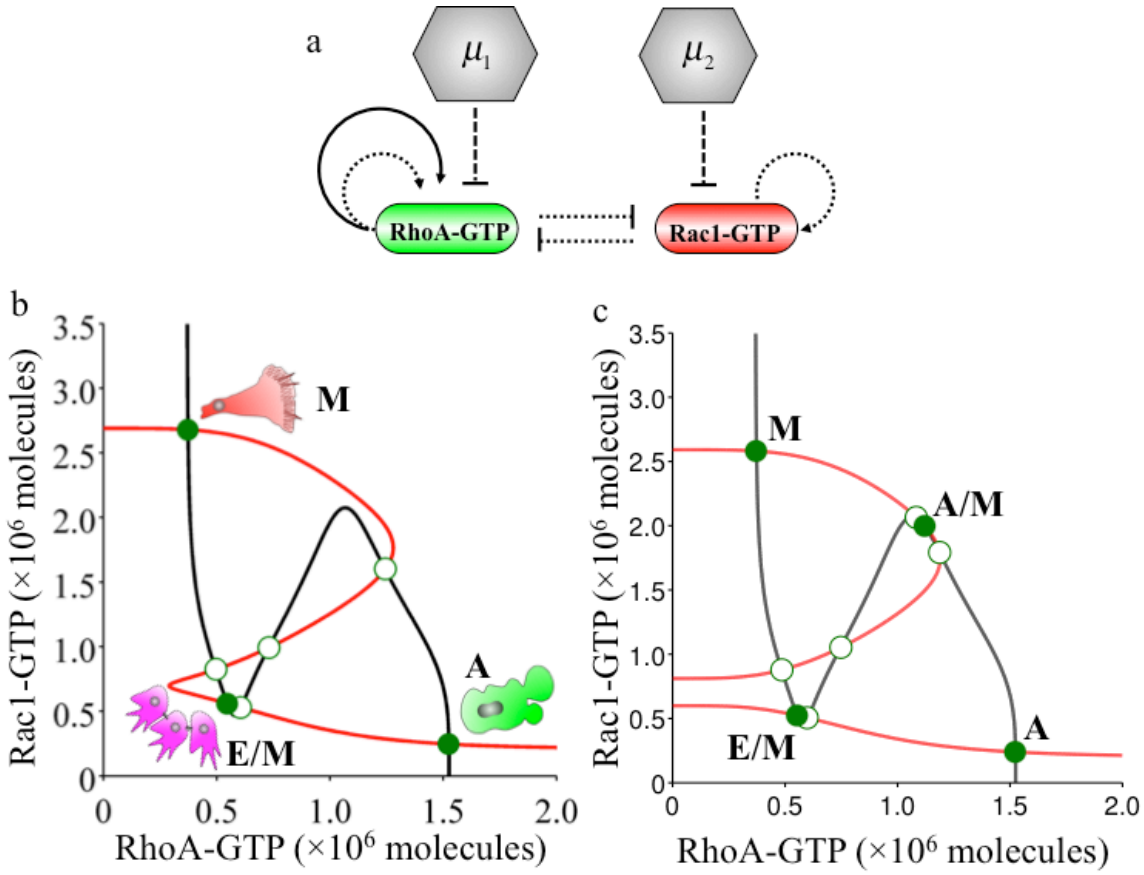


Figure 2.8 The association between the core regulatory circuits for EMT/MET and MAT/AMT.

(a) The effective (reduced) circuit, where μ1 represents the inhibition on RhoA by either miR-34 or miR-200, and μ2 represents similar inhibition on Rac1. A solid arrow denotes activation, and a solid bar denotes repression. A solid line represents transcriptional regulation, a dotted line represents indirect regulations on GTP loading or hydrolysis process via GEFs or GAPs, and a dashed line represents translational inhibition by microRNAs **(b, c)** Dynamical system characteristics of the effective circuit. The plot shows the nullclines and possible steady states corresponding to equation (18). When μ1= μ2 = 100 molecules, the circuit can be tri-stable (M, E/M, A). Red nullcline is for $d[R_c^*]/dt = 0$ and black nullcline is for $d[R_h^*]/dt = 0$. Green solid circles denote the stable steady states, and green hollow circles denote the unstable steady states. Each stable state can be associated with a cell phenotype, as depicted by a cartoon beside them.

The RhoA/Rac1 Circuit Response to microRNA Signals

Next, as the first step towards analyzing the response of the Rac1/RhoA circuit to different levels of the external microRNA signals (μ1 and μ2), we consider μ1 and μ2 to be equal at all times; and that allows us to consider them together as a single external signal denoted by μ (Figure 2.9). Different levels of the signal μ enable different sets of co-existing phenotypes and hence different multi-stable phases, each marked by a distinct color in the figure. At a high level of μ, cells have relatively lower activities of both RhoA and Rac1, and can thereby adopt only the E/M state, i.e. a hybrid E/M phenotype (marked by the pink region). As the level of microRNAs decreases, cells may undergo the transition to adopt a M state or a A state either deterministically or stochastically. With further decrease in the levels of μ, cells can attain relatively high levels of both active RhoA and active Rac1, thereby adopting the A/M state. Our results are consistent

with experiments showing that miR-34 and miR-200 can act as gatekeepers of metastasis and that their decrease leads to collective as well as individual migration[93,94].

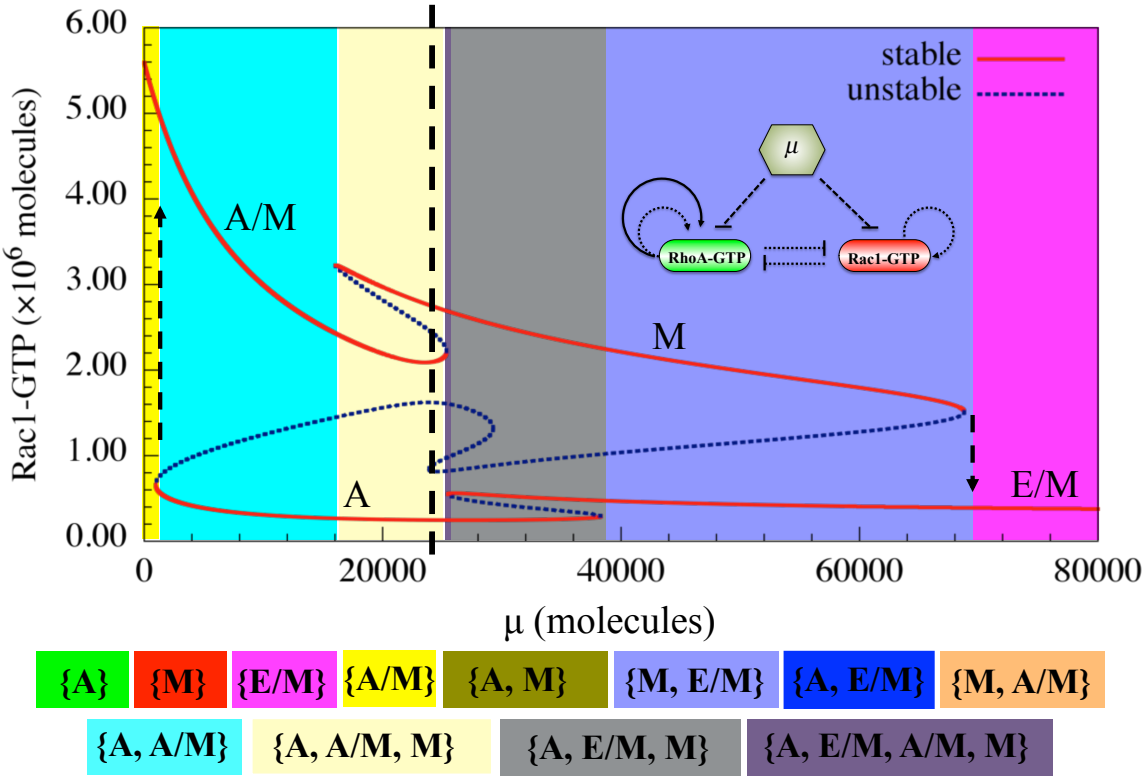


Figure 2.9 Bifurcation of Rac1-GTP protein levels in response to microRNAs (μ) signal regulating the translation of Rac1 and RhoA.

Red solid lines mark the stable states, and blue dashed lines represent unstable states. The bifurcation illustrates the possible co-existence (for some range of the signal) of four states (E/M, M, A/M, and A). The corresponding bifurcation of RhoA-GTP protein levels is shown in Figure S1.5. The co-existence of different phenotypes is associated with a multi-stable phase, highlighted by different background colors (legends given at the bottom). This bifurcation also indicates the possible state transitions when signal μ changes. Starting from the A state (the red line at bottom left part of the diagram), the system could undergo a transition to the A/M state (the red line at top left part of the diagram) when μ signal decreases. The transition is indicated by the dashed upward arrow at the boundary of the phase {A, A/M} (cyan region) and {A/M} (yellow region). Similarly, starting from M state, the increasing μ signal can induce the transition from the

M state to the E/M state, as indicated by the downward arrow at the boundary of the phase {E/M} (pink area) and {E/M, M} (light blue region).

Two-parameter Bifurcation Diagram

Next, we analyze the response of the circuit when the two microRNA signals (μ1 and μ2) can change independently (Figure 2.10). We considered the initial condition such that the cells can adopt any of the three individually migrating phenotypes – A, M, and A/M (marked by black dashed line in Figure 2.9 where μ = 25,000 molecules). At different levels of these two signals (μ1 and μ2), one or several different states or phenotypes can co-exist to form different phases. More specifically, the possible phases are: 1. Phases with only one state – {A}, {M} and {E/M}. 2. Phases in which two states can coexist – {A, E/M}, {M, E/M} and {A, M}. 3. A phase in which three states can coexist – {A, M, E/M} and {A, M, A/M}. 4. A phase in which all four states can coexist – {A, M, E/M, A/M}. These various phases indicate the plasticity of cell migration phenotypes as driven by different combinations of microRNAs (μ1 and μ2 signals).

Depending on how μ1 and μ2 change temporally, the cells follow different trajectories in this phase diagram and thus go through different phenotypic transitions as is illustrated in Figure 2.10b–d. When both the microRNAs – one inhibiting RhoA and the one inhibiting Rac1 – are decreased at the same rate, we see the co-existence of all three solitary migration phenotypes—M, A/M and A phenotypes, i.e. plasticity of cell migration phenotypes is quite rich (Figure 2.10). However, if only one of the microRNAs decreases, cells finally adopt either A or M phenotype (Figure 2.10).

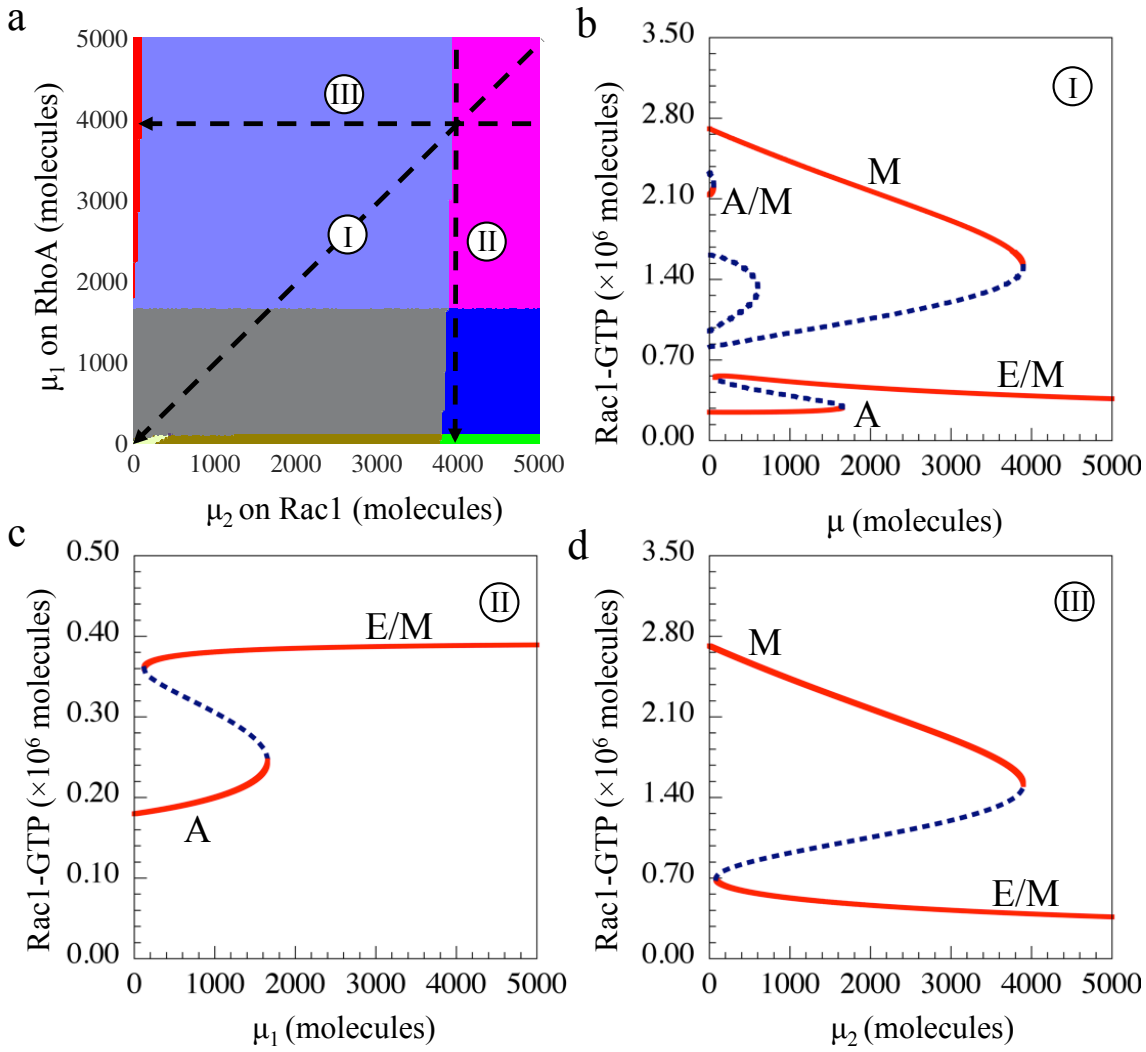


Figure 2.10 The circuit response to the input signals μ1 and μ2.

(a) Two-parameter bifurcation phase diagram, where the two bifurcation parameters are the two input signals μ1 (y-axis) and μ2 (x-axis). Each phase, marked by a distinct color (shown in bottom), corresponds to a different combination of coexisting phenotypes (Phase plane diagrams for each phase are showed in Figure S1.6). **(b)** One-parameter bifurcation diagram when both μ1 and μ2 change simultaneously at the same rate, allowing them to merge into one parameter (here it is referred to as μ), as shown by the trajectory I in the two-parameter bifurcation phase diagram. **(c)** One-parameter bifurcation diagram for the circuit driven by varying levels of μ1 for a fixed μ2 = 4,000

molecules, corresponding to the trajectory II in the two-parameter bifurcation phase diagram. **(d)** One-parameter bifurcation diagram for the circuit driven by varying levels of μ2 at a fixed μ1 = 4,000 molecules, corresponding to the trajectory III in the two-parameter bifurcation phase diagram.

Temporal Dynamics

We demonstrate the temporal dynamics of a phenotypic transition of a cell from collective migration (E/M phenotype) to individual migration (A, M and A/M phenotypes). In the E/M phenotype, cells have high levels of microRNAs (μ1 and μ2). When these signals decrease, cells gradually gain the ability to migrate and can eventually start moving individually in one of the solitary migration phenotypes—A, M and A/M. Notably, in different biological contexts, the two microRNAs (the one inhibiting RhoA (μ1) and the other inhibiting Rac1 (μ2)) may decrease at different rates, thus leading to a difference in the dynamics of phenotypic transition. If the microRNA inhibiting RhoA (μ1) decreases faster than the microRNA inhibiting Rac1 (μ2) (Figure 2.11), the cells transit from the E/M phenotype to A phenotype, i.e. they undergo direct collective to amoeboid transition (CAT) as observed in fibrosarcoma cells[95]. Similar behavior is observed when both microRNAs decrease at same rate (Figure 2.11). In contrast, when (μ2) decreases faster (Figure 2.11), E/M to M phenotype transition may occur (Figure 2.11), i.e. the cells undergo a complete EMT.

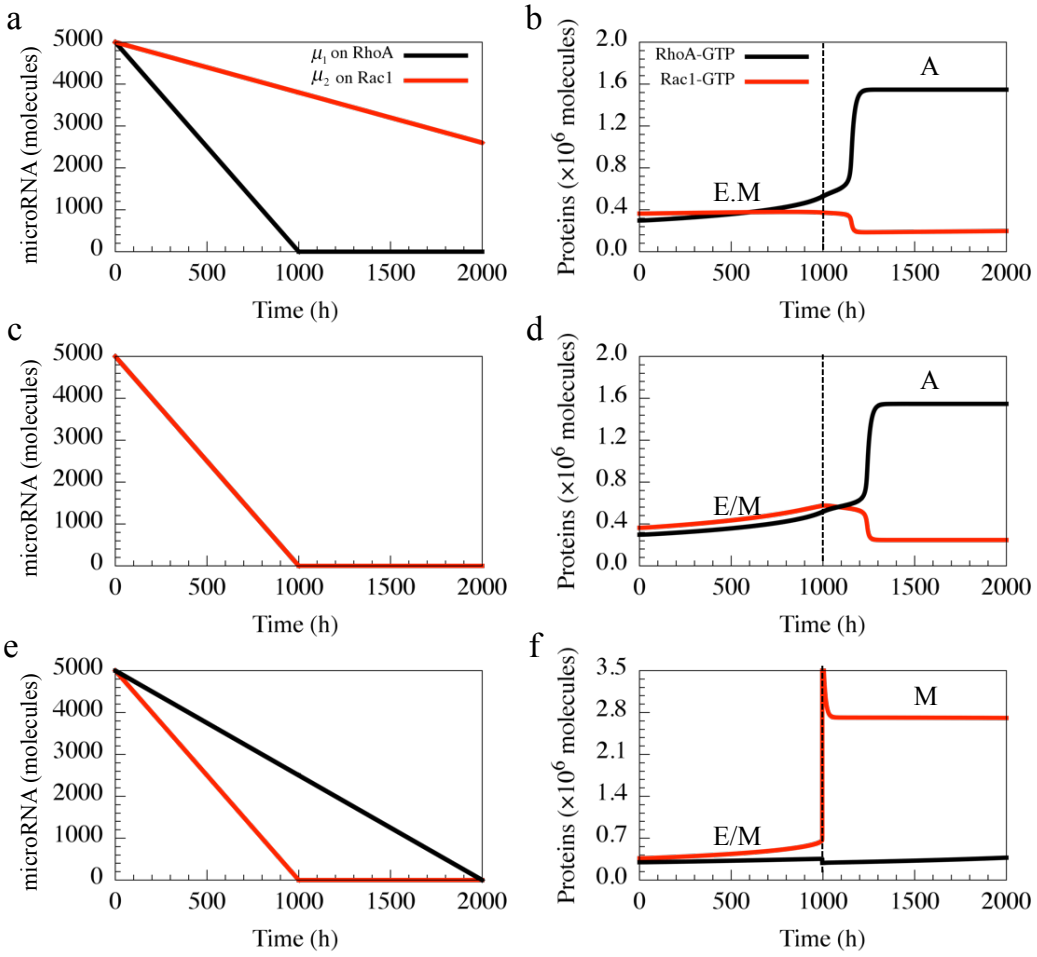


Figure 2.11 Temporal dynamics of the circuit in response to different rates at which μ1 and μ2 decrease.

When μ1 decreases faster than μ2 **(a)**, the cell undergoes a transition from the E/M to the A state **(b)**. When μ1 and μ2 decrease at the same rate **(c)**, cell still switches from the E/M to the M state **(d)**. When μ1 decreases slower than μ2 **(e)**, the cell switches to the E/M state from the M state **(f)**.

Effective Landscape

To understand the relative stability of the co-existing stable states (phenotypes), we construct the effective landscape by considering the biological noise that can be present due to fluctuations in gene expression. Noise can originate from many sources such as birth/death of species and binding/unbinding of proteins, and can induce switches among its stable steady states spontaneously without the action of any external signal. The effective potentials of the landscape are defined as the negative logarithm of the probability (P) of each state ($-\ln(P)$)[96,97]. Therefore, the more common (or frequently adopted) steady state (or phenotype) would have a lower effective potential as shown by blue color. The larger the blue area surrounds a steady state, the more frequently the state or the phenotype is observed.

Here, we have simulated Langevin dynamics of our model by adding Gaussian white noise (see the method section for details) for three cases (Figure 2.12), each of which corresponds to a different level of μ. Without this inhibition signal (μ), four steady states (A, M, A/M, E/M) can be detected. Notably, without the effect of noise, i.e. in the deterministic analysis, the system had three stable states (A, M, A/M) (Figure 2.8), but noise can induce transitions to another steady stable state (E/M). The states representing the A, E/M and M phenotypes are relatively more frequently observed than the state representing the state of the A/M phenotype, as indicated by the higher effective potentials for the A/M basin (Figure 2.12). Also, we observed more transitions between the A and E/M or M and E/M, rather than the other possible transitions (Figure S1.8a). However, with increasing levels of μ, the A/M state disappears and the system could still

stay in the A, M or E/M states, but the probability of staying in the E/M state increases while that for the A and the M states decreases (Figure S1.8b). For even higher levels of μ, the transitions from the collective to the solitary migration is largely inhibited (Figure S1.8) therefore indicating that microRNAs can potentially stabilize the E/M phenotype against the biological noise.

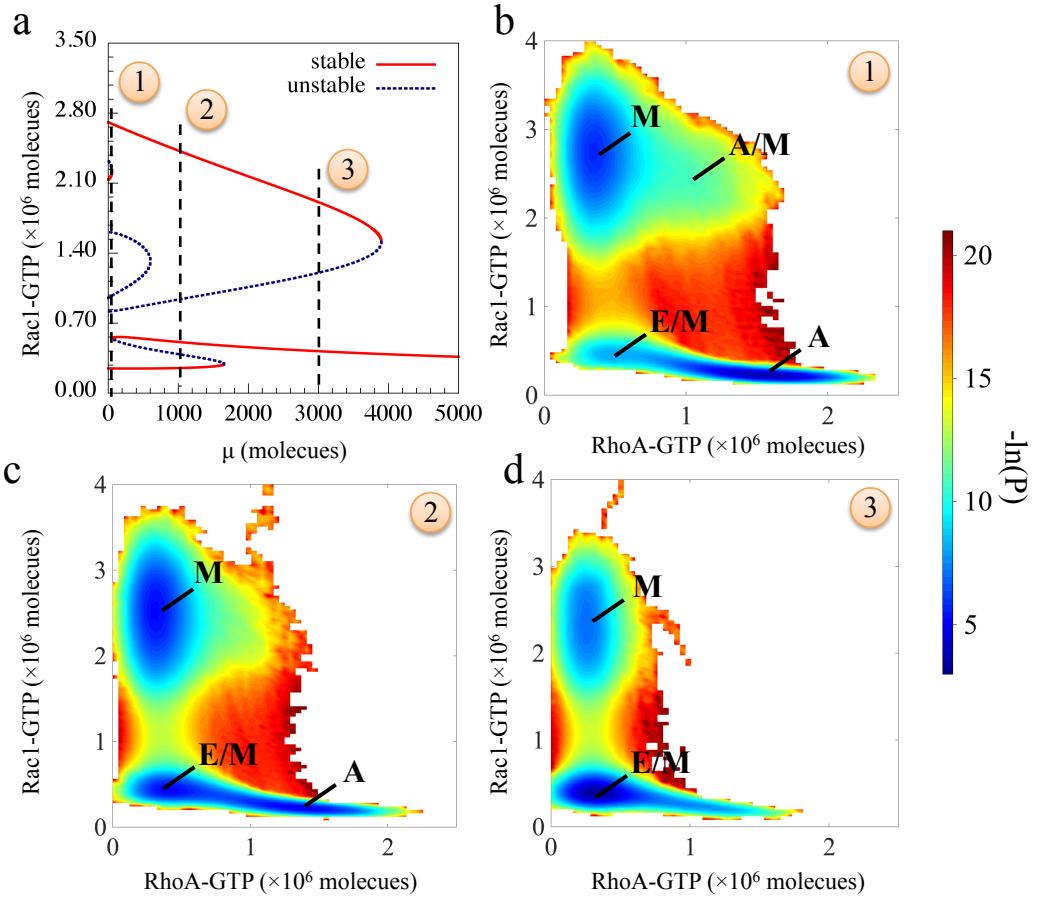


Figure 2.12 Effective landscapes of the circuit with Gaussian white noise.

(a) Bifurcation diagram of the stable Rac1-GTP levels in response to different level of μ. Three different levels of μ are highlighted, as they are the conditions to calculate the effective landscape values (-ln(P)) by Langevin simulation. **(b-d)** show effective

landscapes for (b): $\mu = 0$, (c): $\mu = 1000$, and (d) $\mu = 3000$ (all units in molecules). The basins associated with each state is labeled for each case. The larger the blue region surrounding a particular steady state, the more frequent the state is.

The switch responds to the input signals from c-Met pathway at different microRNA levels

Until now, we have considered the effect of microRNAs on the AMT/MAT circuit in the absence of any EMT-inducing signal. Here, we analyze the response of the AMT/MAT switch or circuit in the presence of an EMT-inducing signal – the HGF/c-Met pathway. c-MET is a tyrosine kinase receptor that is encoded by an oncogene, and is activated when the signaling molecule HGF binds to it[47]. The downstream effectors of the c-MET/HGF pathway include Grb2 and Gab1 that can regulate cell migration phenotypes by activating RhoA or Rac1[98–102] respectively.

To consider the regulatory functions of all these signals combined together, we first calculated the two-parameter bifurcation diagrams (Figure 2.13) when the Rac1/RhoA regulatory circuit is driven by two of these signals – (a) μ and Grb2, and (b) μ and Gab1 (Figure S1.9). As expected, high levels of microRNAs suppress the stimulation of solitary migration by Grb2 and Gab1 signals and therefore retain the cells in the hybrid E/M phenotype. When the level of μ or microRNAs is further decreased, high Grb2 signal can induce the cell to undergo complete EMT to attain the M phenotype, while the regulatory function of Gab1 signal depends on the level of signal μ. Gab1 induces the cells to the M phenotype at the intermediate level of μ, but to A phenotype at the low level of μ. Therefore, the regulatory functions of Grb2 and Gab1 in

inducing cell migration depend on the amount of microRNAs (miR-34 and miR-200), and their functions are more distinguished when the microRNAs are at lower levels, such that Grb2 leads to an M phenotype while Gab1 leads to an A phenotype.

Next, we have calculated the two-parameter bifurcation diagrams of Rac1/RhoA circuit driven by Grb2 and Gab1 signals for various values of μ (Figure 2.13). For the high level of μ, the expression of Rac1 and RhoA is strongly inhibited, thus the plasticity of cell migration driven by Grb2 and Gab1 is limited only to {E/M} and {M} phases, indicating that only EMT/MET could occur (the 1st-2nd panels to the left, Figure 2.13a). However, with a decrease in the levels of μ, cells can also adopt the A state (the 3rd panel to the left, Figure 2.13a); and this phenotypic plasticity for cells only increases with the decrease in the signal μ. At extremely low levels of μ, the cells can adopt any of the three solitary migration phenotypes – A, M, or A/M – depending on the Gab1 and Grb2 levels (the 4th-5th panels to the left Figure 2.13a). Again, these results are consistent with the experiments suggesting that the miR-200 and miR-34 inhibit cell migration and consequently metastasis, and their loss can lead the cells to display a spectrum of solitary migration phenotypes[93,94]. Also, we notice that this series of phase diagrams shows an asymmetry for the induction of different solitary migration phenotypes by Gab1 and Grb2 signals at different microRNA levels. Cells in the E/M phenotype may undergo a transition to the M phenotype (i.e. complete EMT) even at high levels of μ, but they can transit to the A or the A/M phenotypes only at low levels of μ. This asymmetry might underlie why different cell lines might prefer different migration phenotypes (or different distributions of migration phenotypes) during metastasis[37].

Also, for a different parameter set where the Rac1/RhoA circuit is more responsive to Grb2 and Gab1 (the threshold for Grb2 and Gab1 signals was reduced (see supplementary information section 2)). Cells can attain the {A} phase for a larger range of parameters even in the presence of high level of μ, indicating a more metastatic behavior (the 1st-2nd panels to the left, Figure 2.13b). When μ is reduced, the plasticity in cell migration is enhanced significantly (the 3rd-5th panels to the left, Figure 2.13b). This difference in the response to the c-MET pathway can be attributed to different cell lines with distinct metastatic potential.

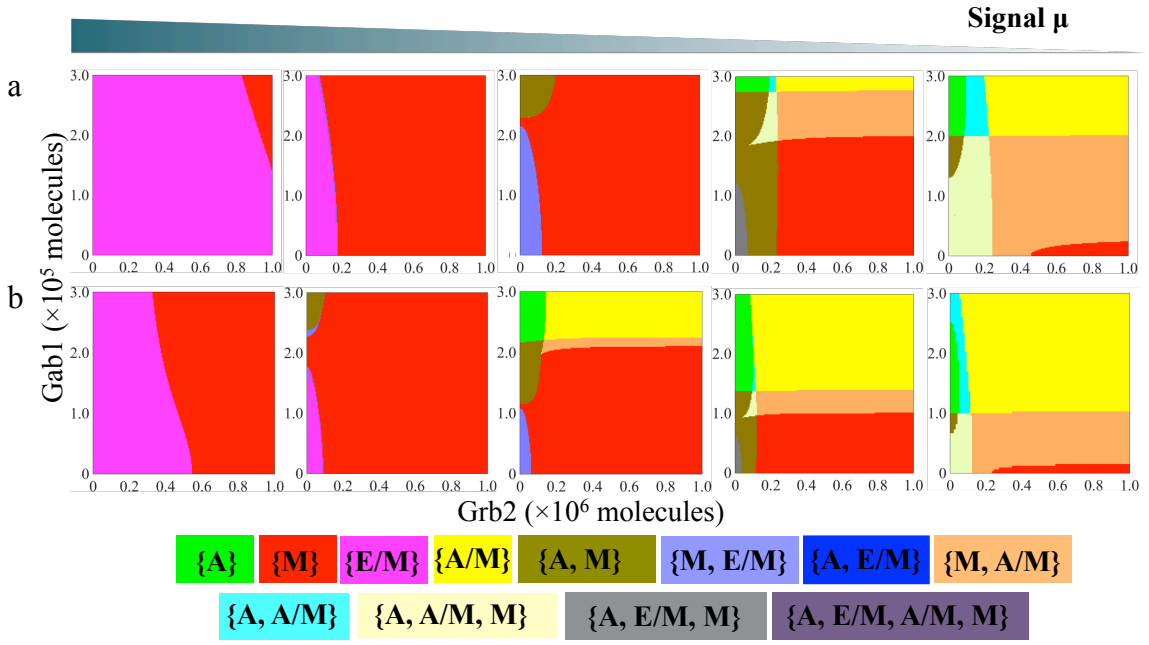


Figure 2.13 The circuit response to Grb1 and Grb2 at different levels of the μ signal.

For both panels a and b, the μ level for each phase diagram decreases from left to right. **(a)** The thresholds of both the Gab1 and Grb2 signals to activate RhoA and Rac1 were set to be 500,000 molecules. At high levels of the microRNAs (first figure of panel a), only two phases are observed – {E/M} and {M}. As the level of μ decreases, the phases with solitary cell migration phenotypes become more common, and many multistable phases

appear. **(b)** The thresholds of both Gab1 and Grb2 to activate RhoA and Rac1 were set to be 250,000 molecules. All other parameters and the μ levels for each case are same as that in panel a. The color scheme for different phases is shown at the bottom.

Phenotype Distribution

Even in the same cell line, cells often have non-heritable phenotypic variability or heterogeneity[88,89]. Such variability can often account for different model parameters for each cell. To capture this heterogeneity, we calculated the population distribution of the phenotypes for the cells with the same levels of microRNAs, but some variability in other model parameters. More specifically, we extended our simulations to a population of 5,000 cells. Each cell has different circuit parameters, which can be ±5% different from the original parameters (details in Method section). In Figure 2.14, we show the percentages of cells in one of the four different possible phenotypes — (A, E/M, M, and A/M) — for different levels of the signal μ. We found that for a high level of signal μ, a significant percentage of cells adopt the E/M phenotype, thereby indicating that the cells with high levels of microRNA are more robust to the parameter variability against maintaining the E/M phenotype.

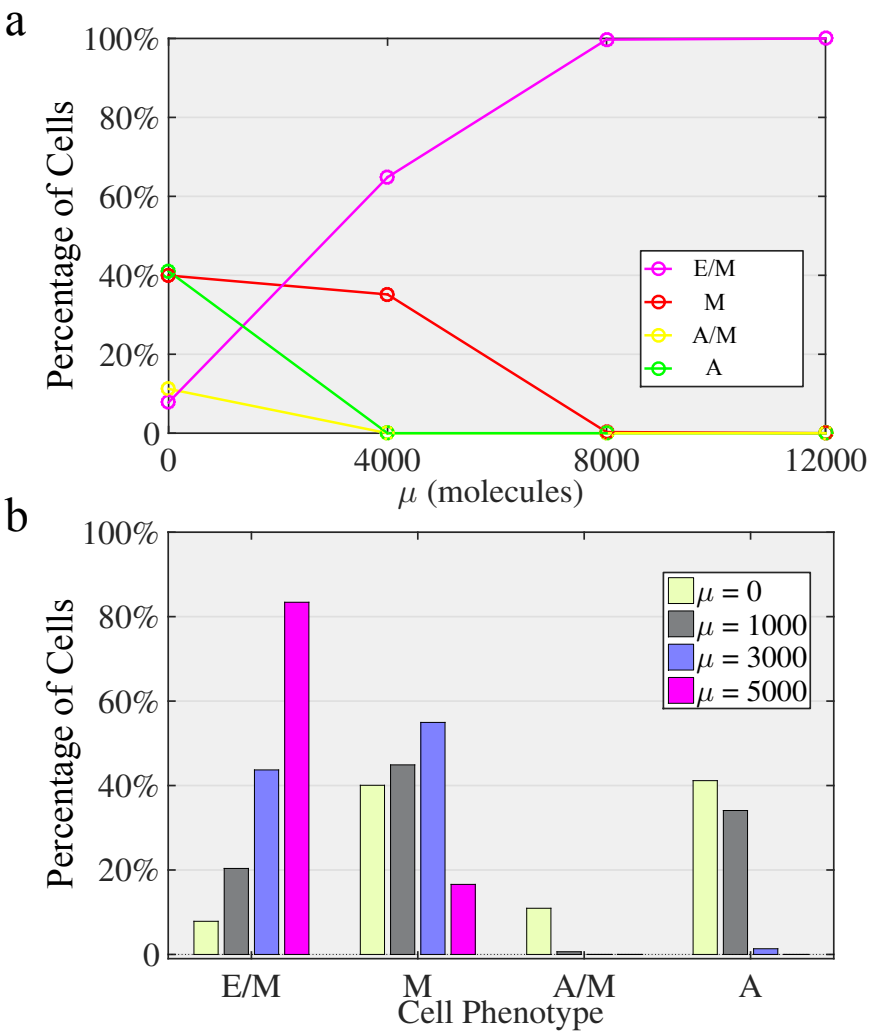


Figure 2.14 Phenotype distribution of a population of cells driven by the signal μ.

(a) The cell parameters are randomly distributed over ±5% relative to the original parameters. The y-axis denotes the percentage of cells corresponding to a specific phenotype. The color of each line represents a phenotype. Under the different levels of μ, a population of cells has different distributions for cell phenotypes (percentage of cells shown in y-axis). **(b)** At different level of signal μ, the original phase of the cells are colored according to the color definition in Figure 2.9. At high levels of the signal μ, the cells are highly likely to stay in the E/M phenotype.

2.5. Discussion

In this study, we first showed that the RhoA/Rac1 circuit could perform as a three-way switch allowing three stable steady states – (high RhoA-GTP, low Rac1-GTP), (low RhoA-GTP, high Rac1-GTP), and (high RhoA-GTP, high Rac1-GTP). These three states can be associated with the A, the M and the hybrid A/M phenotypes respectively, as supported by several experimental evidences. For example, amoeboid cells have been shown to have high actomyosin contractility due to high levels of RhoA-GTP[43]; mesenchymal cells have high actin polymerization due to high levels of Rac1-GTP[43]; the hybrid A/M phenotype displays mixed amoeboid and mesenchymal characteristics, suggesting a comparable level of both actin polymerization and actomyosin contractility due to high levels of both Rac1-GTP and RhoA-GTP[31,35,36]. Notably, similar 'hybrid' manifestations of mixed amoeboid and mesenchymal traits may not be unique to cancer, but can also be observed in other instances, such as neutrophils and leukocytes[76–78].

When incorporating the microRNA signaling on the RhoA/Rac1 circuit, we observed an additional stable state (0, 0), when cells of which have relatively low levels of both RhoA-GTP and Rac1-GTP as compared to the other states (M, A/M, and A). We propose to associate this new state with the hybrid E/M phenotype. Although there is no direct quantitative measurement yet of the active levels of RhoA-GTP and Rac1-GTP in the hybrid E/M cells, our association of the (0, 0) state with hybrid E/M phenotype is consistent with the following experiments. First, moderate levels of both active Rac1 and RhoA have been reported to promote wound healing[91], a typical case of the collective migration of the hybrid E/M cells[27]. Second, a moderate level of active RhoA has been

shown to induce the hybrid E/M phenotype for the human colon adenocarcinoma cell[90], while both dominant-negative and constitutively active Rac1 has been reported to damage the collective migration of boarder cells during Drosophila early development[92]. Both evidences support the relatively low activity of Rac1 and RhoA in collectively migrating hybrid E/M cells as compared with the solitarily migrating cells in A, M and A/M phenotypes.

One limitation of our current model is the exclusion of the epithelial (E) phenotype from our analysis because this model is too simple to explain the elusive dynamics of active RhoA and Rac1 in epithelial tissue establishment and maintenance. Epithelial cells are reported to have higher levels of microRNAs (miR-200 and miR-34) than in the hybrid E/M cells[40,103]. Therefore, according to our model, these cells are likely to have even lower levels of active RhoA and Rac1 than those in hybrid E/M cells. On the other hand, experiments suggest that the active levels of RhoA and Rac1 vary significantly in different stages of epithelial establishment[104]. For instance, at the initial stage of epithelialization, de novo cell-cell adhesion and adherent junctions expansion require the activity of Rac1 and RhoA respectively. But when EMT is induced, levels of active Rac1 decrease at first and then increase again[105]. Therefore, the dynamics of Rac1/RhoA activity need to be fine-tuned either spatially and/or temporally in order to maintain the E-cadherin-mediated cell-cell adhesion. But we currently have not yet modeled the spatial heterogeneity of the two GTPases, RhoA and Rac1, as needed to allow investigation of the role of spatiotemporal dynamics of these proteins in determining the changes in cell shape during epithelialization and transitions between different phenotypes. Besides, to understand the complex temporal dynamics of the

activity of Rac1 and RhoA, additional models should incorporate the interplay among the key proteins in the regulatory circuit such as E-cadherin, N-cadherin, and the Rho GTPases.

From the perspective of dynamical systems, we found that the Rac1/RhoA regulatory circuit can behave as a multi-stable switch in the regulation of different phenotypes. Such multi-stability is a hallmark of self-activating toggle switches (SATS), i.e. mutually inhibitory feedback loop with self-activations on both elements of the loop[106] as seen in RhoA/Rac1 circuit (Figure 2.2). Similar multi-stability is also seen in circuits governing Cancer Stem Cells (CSCs)[107,108] and in circuits governing cell-cell communication[109,110], and can often be used by the cancer cells to adapt to their rapidly changing microenvironments during metastasis[28]. Restricting such cellular plasticity as enabled by these multistable systems can inhibit possible co-operation between different subpopulations and hence hamper tumor metastasis[111–114].

To conclude, our model provides a better understanding of the plasticity of cell migration and its regulation by external signals. Furthermore, it paves a promising way to understand how c-MET pathway is involved in carcinoma metastasis. Given the large number of current attempts to therapeutically target the c-MET pathway, understanding this relationship can provide some useful non-intuitive insights for therapeutic interventions to prevent metastasis.

We also present the first tractable framework towards understanding the transitions among the collective migration phenotype and the solitary migration phenotypes. We found that the microRNAs, miR-34 and miR-200, govern various

phenotypic transitions and can also mediate the effect of other signaling pathways such as Grb2/Gab1 signals on such transitions. Our framework can be further extended by incorporating many other extracellular signals to explore their impact on cellular plasticity and migration, and can possibly provide important insights into cancer therapies that target cell migration during metastasis.

Chapter 3

Development of Random Circuit Perturbation (RACIPE) method to interrogate the topological robustness of gene regulatory circuits

Cells are able to robustly carry out their essential biological functions, possibly because of multiple layers of tight regulation via complex, yet well-designed, gene regulatory networks involving a substantial number of genes. State-of-the-art genomics technology has enabled the mapping of these large gene networks, yet it remains a tremendous challenge to elucidate their design principles and the regulatory mechanisms underlying their biological functions such as signal processing and decision making. One of the key barriers is the absence of accurate kinetics for the regulatory interactions, especially from in vivo experiments. To this end, we have developed a new computational modeling method, Random Circuit Perturbation (RACIPE), to explore the dynamic behavior of gene regulatory circuits without the requirement of detailed kinetic

parameters. RACIPE takes a network topology as the input, and generates an unbiased ensemble of models with varying kinetic parameters. Each model is subjected to simulation, followed by statistical analysis for the ensemble. We tested RACIPE on several gene circuits, and found that the predicted gene expression patterns from the models converge to experimentally observed gene state clusters. We expect RACIPE to be a powerful method to identify the role of network topology in determining network operating principles.

3.1. Introduction

State-of-the-art molecular profiling techniques[115–118] have enabled the construction or inference of large gene regulatory networks underlying certain cellular functions, such as cell differentiation[119,120] and circadian rhythm[121,122]. However, it remains a challenge to understand the operating principles of these regulatory networks and how they can robustly perform their tasks, a prerequisite for cell survival. Mathematical and computational systems biology approaches are often applied to quantitatively model the dynamic behaviors of a network[123–134]. Yet, quantitative simulations of network dynamics are usually limited due to several reasons. First, a proposed network might contain inaccurate or missing regulatory genes or links, and modeling an incomplete network might produce inaccurate predictions. Second, kinetic parameters for each gene and regulatory interaction, which are usually required for quantitative analyses, are difficult to be obtained directly for all of them from *in vivo* experiments[135]. To deal with this problem, network parameters are either inferred from

existing data [136,137] or educated guesses, an approach which could be time-consuming and error-prone. This approach is hard to extend to very large gene networks due to their complexity.

Alternatively, a bottom-up strategy has been widely used to study the regulatory mechanisms of cellular functions. First, one performs a comprehensive analysis and integration of experimental evidence for the essential regulatory interactions in order to construct a core regulatory circuit, typically composed of only a small set of essential genes. The core gene circuit is then modeled either by deterministic or stochastic approaches with a particular set of parameters inferred from the literature. Due to the reduced size of the systems and the inclusion of data derived directly from the literature, the bottom-up approach suffers less from the above-mentioned issues. Examples of the bottom-up approach include the modeling of biological process such as Epithelial-to-Mesenchymal transition (EMT)[12,13,40], cell cycles[14,15], and circuit design in synthetic biology, such as genetic toggle switch[3], and repressilator[4].

Due to the success of these and other circuit-based modeling studies, we hypothesize that a core circuit module should emerge from a complex network and dictate the decision-making process. It is reasonable to assume that a large gene network could be decomposed into a core gene circuit and a peripheral part with the residual genes. The core would then be the driving force for the network dynamics and should be robust against cell-to-cell variability and extrinsic fluctuations in stimuli arising from cell signaling, while the peripheral genes would act to regulate the signaling status for the core circuit and probably also enhance the robustness of the core dynamics by

introducing redundancy. This scale-separation picture is consistent with ideas such as the existence of master regulators and network modularity[138,139].

On the basis of this conceptual framework, we developed a new computational method, named *ra*ndom *ci*rcuit *pe*rturbation (RACIPE), for modeling possible dynamic behaviors that are defined by the topology of a core gene regulatory circuit. In RACIPE, we focus the modeling analysis on the core circuit and regard the effects of the peripheral genes and external signaling as random perturbations to the kinetic parameters. In contrast to traditional modeling methods[140], RACIPE generates an ensemble of mathematical models, each of which has a different set of kinetic parameters representing variations of signaling states, epigenetic states, and genetic backgrounds (including cells with genetic mutations leading to disease). Here we randomize the model parameters by one or two orders of magnitude and utilize a specially designed parameter sampling scheme (details in Methods) to capture the key role of circuit topology. This random field approach allows the inclusion of the contributions from the peripheral genes to the network dynamics and the evaluation of their roles in modulating the functions of the core circuit. From the *in silico* generated data, we apply statistical analysis to identify the most probable features within all of the models, a process which can uncover the most robust functions of the core circuit. It is worth-noting that RACIPE is unique in the way it utilizes perturbation and the integration of statistical tools, compared to the traditional parameter sensitivity analysis[141–145] and the previous studies on random circuit topology[146,147].

In the following, we will first describe in detail the RACIPE method, and then present the results of applying RACIPE to several simple standalone circuit motifs and also coupled toggle switch motifs. In addition, we will show the application of RACIPE to a 22-component network for the decision-making core of the Epithelial-to-Mesenchymal Transition (EMT). We will see that RACIPE is capable of identifying accessible gene states via statistical analysis of the *in silico* generated data, from which we can further decode the design principles and evaluate the robustness of the core gene circuit. We therefore expect RACIPE to be a powerful tool to analyze the dynamic behavior of a gene network and to evaluate the robustness and accuracy of proposed network models.

3.2. Random Circuit Perturbation (RACIPE) Method

We developed a new computational method, namely *ra*ndom *ci*rcuit *pe*rturbation (RACIPE), for modeling a gene network. The procedure of RACIPE is illustrated in Figure 3.1. The input of RACIPE is the topology of the core circuit under study, which can be constructed according to either the literature, interaction databases (e.g. Ingenuity pathway analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity), KEGG[17], GO[18]), or computational methods[19]. From the circuit topology, we establish a set of mathematical equations for the time evolution of the levels of all the genes. We then generate an ensemble of models where the parameters from the rate equations are sampled by a carefully designed randomization procedure (see below for details) so that these kinetic models can capture the behavior of the circuits under

different conditions. Each model is subject to standard analysis to discover possible dynamics of the circuit (Figure 3.1b). The dynamics could converge to a stable steady state, a stable oscillation, or chaotic behavior. To find all possible behaviors for a gene network, we typically choose many different sets of initial conditions (randomly sampled on a logarithmic scale) and numerically solve the rate equations for each case. The procedure is repeated for many times to collect sufficient data for statistical analysis. In particular, for a multi-stable system, this ODE-based method is useful for identifying all the distinct stable states for a multi-stable system. Thus, the RACIPE method can generate a large amount of simulated gene expression data, which can be further analyzed by biostatistics tools (Figure 3.1c). Potentially, RACIPE can also be extended to study oscillatory or adaptive dynamics (Figure S2.1), and is also compatible with other types of modeling methods such as stochastic analysis, but these are out of scope this study. In the following, we will illustrate RACIPE in the context of a multi-stable gene circuit by deterministic analysis. In the following, we will illustrate RACIPE in the context of a multi-stable gene circuit by deterministic analysis.

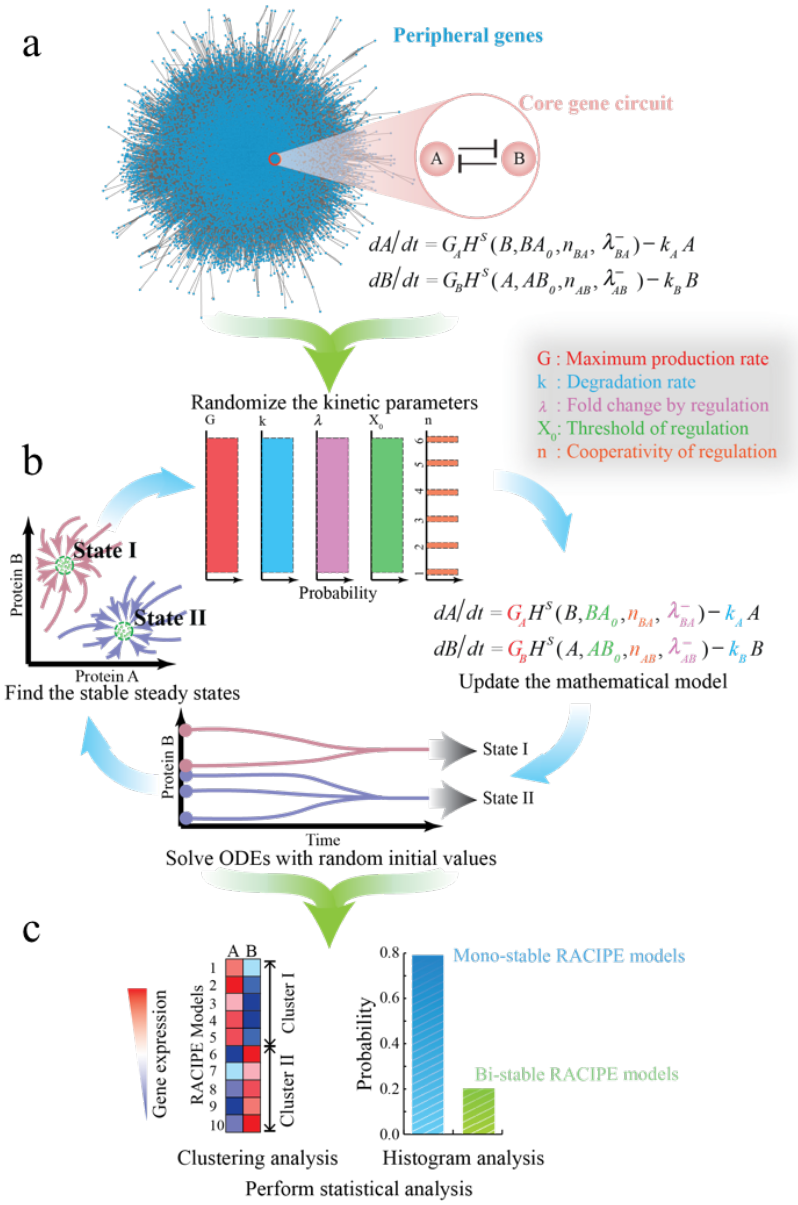


Figure 3.1 Schematics of the random circuit perturbation (RACIPE) method.

(a) The gene regulatory network for a specific cellular function is decomposed into two parts – a core gene circuit modeled by chemical rate equations and the other peripheral genes whose contribution to the network is regarded as random perturbations to the kinetic parameters of the core circuit. **(b)** RACIPE generates an ensemble of models, each of which is simulated by the same rate equations but with randomly sampled kinetic parameters. For each model, multiple runs of simulations are performed, starting from

different initial conditions, to identify all possible stable steady states. **(c)** The *in silico* gene expression data derived from all of the models are subject to statistical analysis

As an example, we start with the deterministic rate equations for a toggle switch circuit (Figure 3.2) with mutually inhibitory genes A and B. The kinetic model takes the form:

$$\begin{aligned}\dot{A} &= g_A H^S(B, BA_0, n_{BA}, \lambda_{BA}^-) - k_A A \\ \dot{B} &= g_B H^S(A, AB_0, n_{AB}, \lambda_{AB}^-) - k_B B\end{aligned}, \quad (19)$$

where A and B represent the expression levels of gene A and B respectively. g_A and g_B are the basal production rates (the production rate for the gene without any regulator bound to the promoter). k_A and k_B are the innate degradation rates. Regulation of gene B expression by A is formulated as a non-linear shifted Hill function ($H^S(A, AB_0, n_{AB}, \lambda_{AB}^-)$), defined as $\lambda_{AB}^- + (1-\lambda_{AB}^-)H^-(A, AB_0, n_{AB})$, where $H^- = 1/(1+(A/AB_0)^{n_{AB}})$ is the inhibitory Hill function, AB_0 is the threshold level for A, n_{AB} is the Hill coefficient of the regulation, λ_{AB}^- is the maximum fold change of the B level caused by the inhibitor A ($\lambda_{AB}^- < 1$). In the case of an activator, the fold change is represented by λ_{AB}^+ ($\lambda_{AB}^+ > 1$). The inhibitory regulation of gene A by gene B can be modeled in an analogous way.

In RACIPE, randomization is performed on all five types of circuit parameters: two of them are associated with each gene, including the basal production rate (g) and the degradation rate (k); and three of them are associated with each regulatory link,

including the maximum fold change of the gene expression level (λ), the threshold level of the regulation (X_0) and the Hill coefficient (n).

Our parametric randomization procedure is designed to ensure that the models can represent all biologically relevant possibilities. In detail, the Hill coefficient n is an integer selected from 1 to 6, and the degradation rate k ranges from 0.1 to 1 (See Table S2.1 for the explanation of the units). Here each parameter is assigned by randomly picking values from either a uniform distribution or some other distributions, for example the Gaussian distribution. The fold change λ^+ ranges from 1 to 100 if the regulatory link is excitatory, while λ^- was varied from 0.01 to 1 if the regulatory link is inhibitory. Note that for the latter case, a probability distribution (e.g. a uniform distribution) is sampled for the inverse of λ^-, i.e. $1/\lambda^-$, instead of λ^- itself. By doing so, we make sure that the mean fold change is about 0.02, instead of ~ 0.5. The choice of such a wide range of λ values ensures the consideration of both strong and weak interactions.

In addition, two assumptions are made in RACIPE to ensure that it generates a representative ensemble of models for a specific circuit topology. First, the maximum production rate of each gene should lie roughly within the same range (from 1 to 100 in this study, see Table S2.1), as the maximum rate is determined by how fastest the transcriptional machinery can work. For a gene regulated by only one activator, the maximum production rate is achieved when the activator is abundant, and thus the basal production rate of the gene $g = G/\lambda^+$. For a gene regulated by only one inhibitor, the maximum rate is achieved in the absence of the inhibitor, i.e. $g = G$. This criterion can

be generalized to genes regulated by multiple regulators. Therefore, in practice, we directly randomize the maximum production rate (G) instead of the basal production rate (g), and then calculate the value of g according to the above criterion. The ranges of these parameters are summarized in details in Table S2.1. The RACIPE randomization procedure allows a gene to have a relative expression ratio of up to 1,000 for two sets of parameters, even when it is not regulated by other genes.

Second, we also assume that, for all the members of the RACIPE model ensemble, each regulatory link in the circuit should have roughly equal chance of being functional or not functional, referred to as the *half-functional* rule. For example, in the case that gene A regulates gene B, all the parameters are selected in such a way that for the RACIPE ensemble, the level of A at the steady states has roughly 50% chance to be above and 50% chance to be below its threshold level. Otherwise, if the threshold level is too large or too small, the regulatory link is either not functional most time or constitutively active, thereby changing the effective circuit topology, and limiting a comprehensive understanding of circuit function (Figure S2.2).

To achieve this, we estimate the range of the threshold levels by a mean-field approximation, and use this range for randomly sample of the threshold parameters. For a regulatory link from gene A (regulator) to gene B (target), the threshold level AB_0 can be estimated as follows. We first estimate the range of expression of gene A without considering any of its regulators. The A level without regulation satisfies

$$\dot{A} = G - kA, \quad (20)$$

By randomizing both G and k by the aforementioned protocol (Table S2.1), we generate an ensemble of random models, from which we obtain the distribution of the steady state levels of gene A (Figure 3.2a). To meet the half-functional rule, the median of the threshold level should be chosen to be the median of this distribution. When gene A is regulated by some other genes (i.e. its upstream regulators), we estimate the median threshold level by considering the inward regulators of A, and assume that the levels of all these regulators (*e.g.* gene B, C *etc.*) follow the same distribution as an isolated gene (top right panels in Figure 3.2a section 2). We set the threshold of every inward regulation to be 0.02M to 1.98M, where M is the median of the distribution of an isolated gene. With all of the information, we can again generate a new ensemble of models, from which we calculate the distribution of gene A (bottom panel in Figure 3.2a section 2) and its median. For every target gene regulated by the gene A, we select the threshold levels of the regulations in the range from 0.02M to 1.98M, where M is the above obtained median level of gene A. The same approach is used to estimate the threshold levels of the other genes. It is worth-noting that this self-consistent strategy works quite well for the cases we have tested (Figure 3.2b) according to the *half-functional* rule.

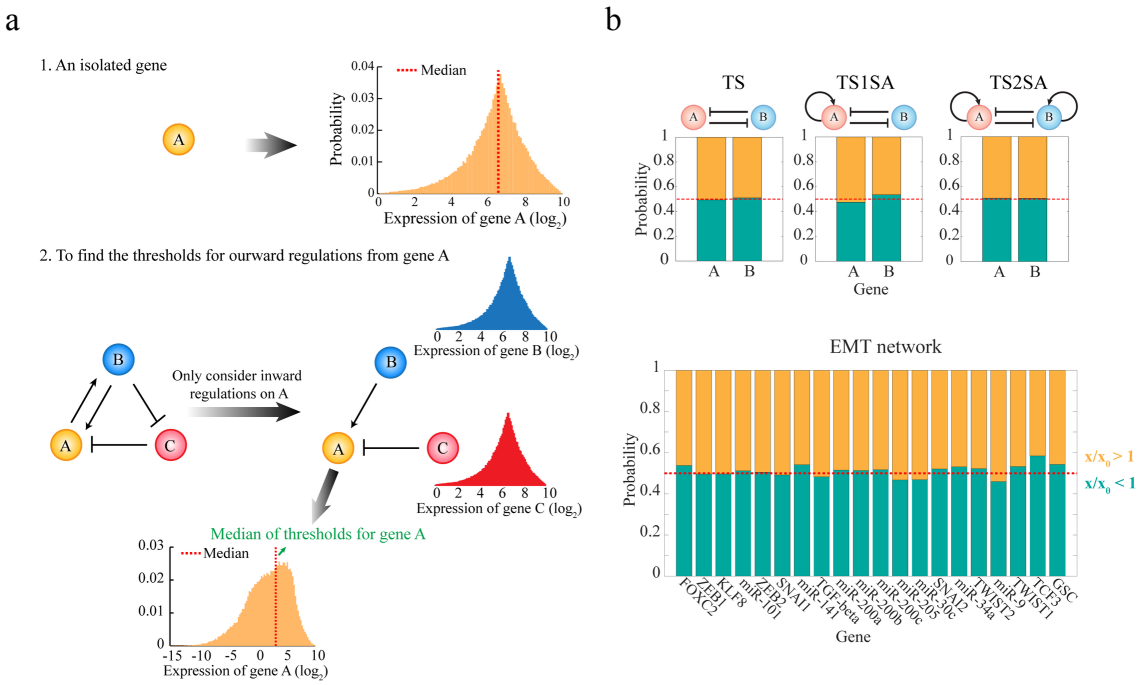


Figure 3.2 Randomization scheme to estimate the ranges of the threshold parameters.

(a) Schematics of the procedure to estimate the ranges of the threshold parameters, so that the level of a regulator has 50% chance to be above or below the threshold level of each regulatory link ("half-functional rule"). First, for a gene A without any regulator, the RACIPE models are generated by randomizing the maximum production rate and the degradation rate according to Table S2.1. The distribution of A level is obtained from the stable steady state solutions of all the RACIPE models (top left panel, yellow histogram). Second, for a gene A in a gene circuit, the distribution of A level is estimated only on the basis of the inward regulatory links (i.e. the B to A activation and the C to A inhibition in the bottom left panel). The distributions of the levels of the inward regulators B and C are assumed to follow the same distributions as a gene without any regulator (bottom left panel, blue and red distribution); the threshold levels for these inward links are chosen randomly from (0.02M to 1.98M), where M is the median of their gene expression distributions. Finally, the distribution of A level is obtained by randomizing all the relevant parameters. That includes the levels of B and C, the strength of the inward regulatory links, the maximum production rate and the degradation rate of the A, and the threshold for any regulatory link starting from A is chosen randomly from (0.02M to 1.98M), where M is the median level of the new distribution of A level (orange in the bottom panel). The same procedure is followed for all other genes. **(b)** Tests on several

simple toggle-switch-like circuit motifs and the Epithelial-to-Mesenchymal Transition (EMT) (Figure 3.6a) circuit show that the "half-functional rule" is approximately satisfied with this randomization scheme. For each RACIPE model, we computed the ratio (x/x_0) of the level of each gene X at each stable steady state (x) and the threshold (x_0) for the outward regulations from gene X. The yellow region shows the probability of $x/x_0 > 1$ for all the RACIPE models, and the green region shows the probability of $x/x_0 < 1$.

3.3. Applications

3.3.1. Simple toggle-switch like circuits

3.3.1.1. Introduction

We first tested RACIPE on several basic toggle-switch-like circuit motifs (Figure 3.3a). These circuit motifs are considered to be some of the main building blocks of gene regulatory networks[148]. A genetic toggle switch (TS), composed of two mutually inhibitory genes, is commonly considered to function as a bi-stable switch - it allows two stable gene states, each of which is characterized by dominant expression of one gene. The TS has been shown to be a central piece of decision-making modules for cell differentiation in several incidences[149–151].

3.3.1.2. Methods

Mathematical models of toggle-switch like circuit motifs

The circuits, including simple toggle-switch-like circuit motifs, coupled toggle-switch (CTS) motifs, and the EMT circuit, were modeled by the deterministic rate equations as follows:

Toggle-switch circuit (TS):

$$\begin{aligned}\dot{A} &= G_A H^S(B, BA_0, n_{BA}, \lambda^-_{BA}) - k_A A \\ \dot{B} &= G_B H^S(A, AB_0, n_{AB}, \lambda^-_{AB}) - k_B B\end{aligned}, \quad (21)$$

Toggle-switch circuit with one-sided self-activation (TS1SA):

$$\begin{aligned}\dot{A} &= G_A H^S(B, BA_0, n_{BA}, \lambda^-_{BA}) - k_A A \\ \dot{B} &= G_B H^S(A, AB_0, n_{AB}, \lambda^-_{AB}) H^S(B, BB_0, n_{BB}, \lambda^+_{BB}) / \lambda^+_{BB} - k_B B\end{aligned}, \quad (22)$$

Toggle-switch circuit with two-sided self-activation (TS2SA):

$$\begin{aligned}\dot{A} &= G_A H^S(B, BA_0, n_{BA}, \lambda^-_{BA}) H^S(A, AA_0, n_{AA}, \lambda^+_{AA}) / \lambda^+_{AA} - k_A A \\ \dot{B} &= G_B H^S(A, AB_0, n_{AB}, \lambda^-_{AB}) H^S(B, BB_0, n_{BB}, \lambda^+_{BB}) / \lambda^+_{BB} - k_B B\end{aligned}, \quad (23)$$

Coupled toggle-switch (CTS) motif with n toggle switches:

$$\begin{aligned}
&\dot{A}_1 = G_{A_1} H^S(B_1, B_1 A_{10}, n_{B_1 A_1}, \lambda^-_{B_1 A_1}) - k_{A_1} A_1 \\
&\dot{A}_i = G_{A_i} H^S(B_i, B_i A_{i0}, n_{B_i A_i}, \lambda^-_{B_i A_i}) H^S(A_{i-1}, A_{i-1} A_{i0}, n_{A_{i-1} A_i}, \lambda^+_{A_{i-1} A_i}) / \lambda^+_{A_{i-1} A_i} - k_{A_i} A_i \qquad (i > 1) \\
&\dot{B}_j = G_{B_j} H^S(A_j, A_j B_{j0}, n_{A_j B_j}, \lambda^-_{A_j B_j}) H^S(B_{j+1}, B_{j+1} B_{j0}, n_{B_{j+1} B_j}, \lambda^+_{B_{j+1} B_j}) / \lambda^+_{B_{j+1} B_j} - k_{B_j} B_j \quad (j < n) \\
&\dot{B}_n = G_{B_n} H^S(A_n, A_n B_{n0}, n_{A_n B_n}, \lambda^-_{A_n B_n}) - k_{B_n} B_n
\end{aligned}, \quad (24)$$

where G_i represents the maximum production rate of gene i, and k_i represents the degradation rate of gene i. For the regulatory link (either activation or inhibition) from gene i to gene j, ij_0 represents the threshold level of gene i, n_{ij} is the Hill coefficient, and λ_{ij} is the maximum fold change of the gene j level caused by gene i. $\lambda_{ij} < 1$ for an inhibitory regulation, referred to as λ_{ij}^-, while $\lambda_{ij} > 1$ for an excitatory regulation, referred to as λ_{ij}^+. For the simple gene circuits (TS, TS1SA, TS2SA), A and B stand for the expression levels of the two transcription factors. The notation for each gene in the CTS motifs is shown in Figure 3.5. Through RACIPE method, we generated 10,000 sets of parameters (10,000 RACIPE models) for each circuit. Each model is solved by the Euler method starting from 1,000 different initial conditions to find out all the distinct stable steady solutions.

Normalization of gene expression data

From an ensemble of the RACIPE models, we collected a large set of gene expression data, very similar to those obtained in experiment. Therefore, we can apply tools for data analysis to these computationally generated data. Here, we processed the computational expression data by a standard normalization method before we performed clustering analysis. The gene expression levels from the models are normalized by first log transformation and standardization, i.e.

$$x_i \to \frac{\log_2(x_i) - \overline{\log_2(x_i)}}{\sigma(\log_2(x_i))}, \quad (25)$$

Clustering analysis

We performed cluster analysis on the normalized gene expression data from the RACIPE models. Each column represents a gene, and each row represents a stable steady state of the circuit for a particular RACIPE model. For RACIPE models with more than one stable state, the gene expression profiles for all of the stable states were listed in multiple rows. We applied average linkage hierarchical clustering analysis using Euclidean distance by Cluster 3.0[152] and the results were visualized by JavaTreeview[153]. The cutoff distance was chosen interactively so that major clusters can be identified and each cluster has a distinct gene expression pattern. These clusters correspond to different gene states for the circuit. Principal component analysis was performed on the same data by using "pca" function in Matlab/2014b. The major gene clusters can be readily recognizable from the probability density map projected onto the first two principal component axes.

3.3.1.3. Results

For TS motif, we observed that about 20% of models allow two coexisting stable steady states (bi-stability), while the remainder allow only one steady state (mono-stability). The observation that only a small fraction of TS models work as a bi-stable system is consistent with a previous study[146]. Next, we tested RACIPE on a toggle switch with an extra excitatory auto-regulatory link acting on only one of the genes (a toggle switch with one-sided self-activation, or TS1SA). The circuit motif now has ~ 50% chances of being bi-stable, much larger than the original TS motif. Interestingly, TS1SA also has ~1% chance of having three co-existing stable steady states (tri-

stability), so it can potentially act as a three-way switch[148]. Hence, the RACIPE analysis suggests that TS1SA is more robust than TS for functioning as a switch. Moreover, adding excitatory auto-regulatory links on both sides of the TS motif (TS2SA) further increases the likelihood of bi-stability to ~60%, and meanwhile dramatically increases the likelihood of tri-stability to ~13%. This suggests that TS2SA has more of an ability than these other motifs to function as a three-way switch. Indeed, TS2SA has been proposed to be the core decision-making motif for several cell differentiation processes, and many of these processes exhibit multi-stability[149,150]. Thus, statistical analysis of the ensemble of random models generated by RACIPE can identify the most robust features of a circuit motif.

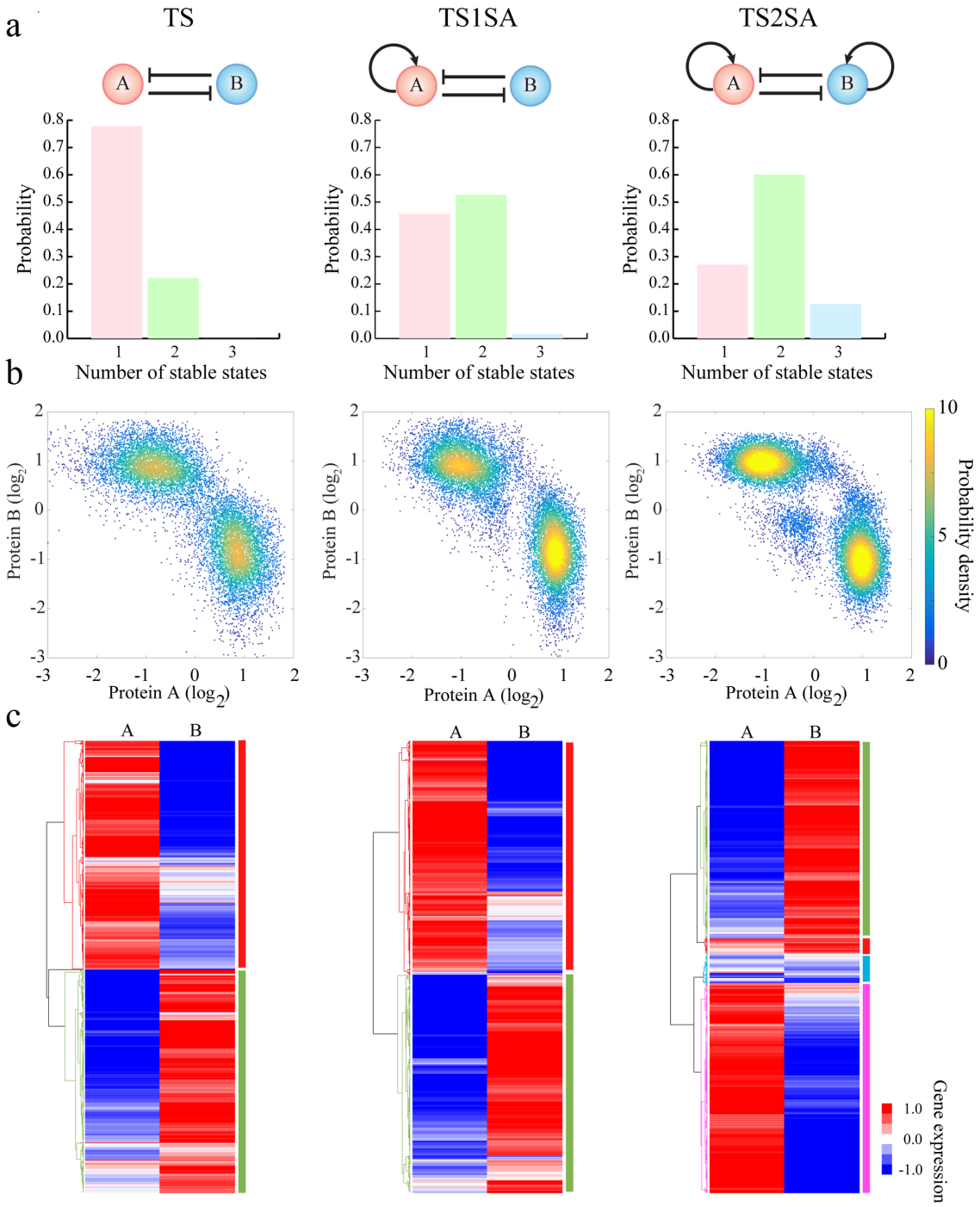


Figure 3.3 RACIPE identifies robust features of toggle-switch-like motifs.

RACIPE was tested on three circuits – a simple toggle-switch (TS, top left) which consists of genes A and B that mutually inhibit each other (solid lines and bars), a toggle-

switch with one-sided self-activation (TS1SA) which has an additional self-activation link on gene B, and a toggle-switch with two-sided self-activation (TS2SA) which has additional self-activation links on both genes. **(a)** Probability distributions of the number of stable steady states for each circuit. **(b)** Probability density maps of the gene expression data from all the RACIPE models. Each point represents a stable steady state from a model. For any RACIPE model with multiple stable steady states, all of them are shown in the plot. **(c)** Average linkage hierarchical clustering analysis of the gene expression data from all the RACIPE models using the Euclidean distance. Each column corresponds to a gene, while each row corresponds to a stable steady state from a model. The analysis shows that the gene expression data could be clustered into distinct groups, each of which is associated with a gene state, as highlighted by different colors on the right of the heatmaps.

Another way to utilize RACIPE is to evaluate the possible gene expression patterns of a circuit motif. We can construct a large set of *in silico* gene expression data, consisting of the gene expression levels of the circuit at every stable steady state for each RACIPE model. In the dataset, the column corresponds to the genes and the rows corresponds to stable steady states. For a RACIPE model with multiple stable steady states, we enter the data in multiple rows. The expression dataset takes a form similar to typical experimental microarray data, and so it can be analyzed using common bioinformatics tools. For each of the above two-gene cases, we visualized the expression data by a scatter plot of the levels of the two genes (Figure 3.3b). Surprisingly, despite large variations in the circuit parameters across the RAICPE model ensemble, the expression data points converge quite well into several robust clusters. For example, the TS circuit data has two distinct clusters, where one has a high expression of gene A but a low expression of gene B and the other vice versa. The TS2SA circuit has not only the above two clusters but also an additional cluster with intermediate expression of both genes. These patterns have also been observed in previous experimental[3] and theoretical[148,149,154] studies of the same circuits. Interestingly, if we only include

models with a fixed number of stable states (e.g. restrict the ensemble to mono-stable models, or bi-stable models), a similar pattern of clusters can still be observed (Figure S2.3). These clusters represent distinct patterns of gene expression that the circuit can support, so we will refer to these clusters as "gene states". These gene states are robust against large perturbations of circuit parameters because the circuit topology restricts possible gene expression patterns. RACIPE in a sense takes advantage of this feature to interrogate the circuit so that we can unbiasedly identify the robust gene states. Since these states may be associated with different cell phenotypes during cell differentiation or cellular decision-making processes, RACIPE can be especially helpful in understanding the regulatory roles of the circuit during transitions among different states.

These simple cases demonstrate the effectiveness of RACIPE in revealing generic properties of circuit motifs. Recall that our basic hypothesis is that the dynamic behaviors of a circuit should be mainly determined by circuit topology, rather than a specific set of parameters. The rich amount of gene expression data generated by RACIPE allows the application of statistical learning methods for the discovery of these robust features. For example, as shown in Figure 3.3c, we applied unsupervised hierarchical clustering analysis (HCA) to the RACIPE gene expression data, and again we identified similar gene state clusters for each circuit.

Notably, the predictions of these gene states by RACIPE should be robust against different sampling distributions and different ranges of kinetic parameters. To verify this, we tested on the TS circuit versions of RACIPE created with three different distributions (uniform, Gaussian and exponential distribution) and three different ranges of parameters

(Figure 3.4). Even though the precise shape of gene states appears to be slightly different for the different cases, the number and the locations of these gene states are consistent (Figure 3.4). For the cases with exponential distribution, when the range of the parameters is reduced, the mean decreases as well; therefore, the two gene states become closer (Figure 3.4). We also found that the changes of the parameter range still result in similar gene states (Figure S2.4 and S2.5).

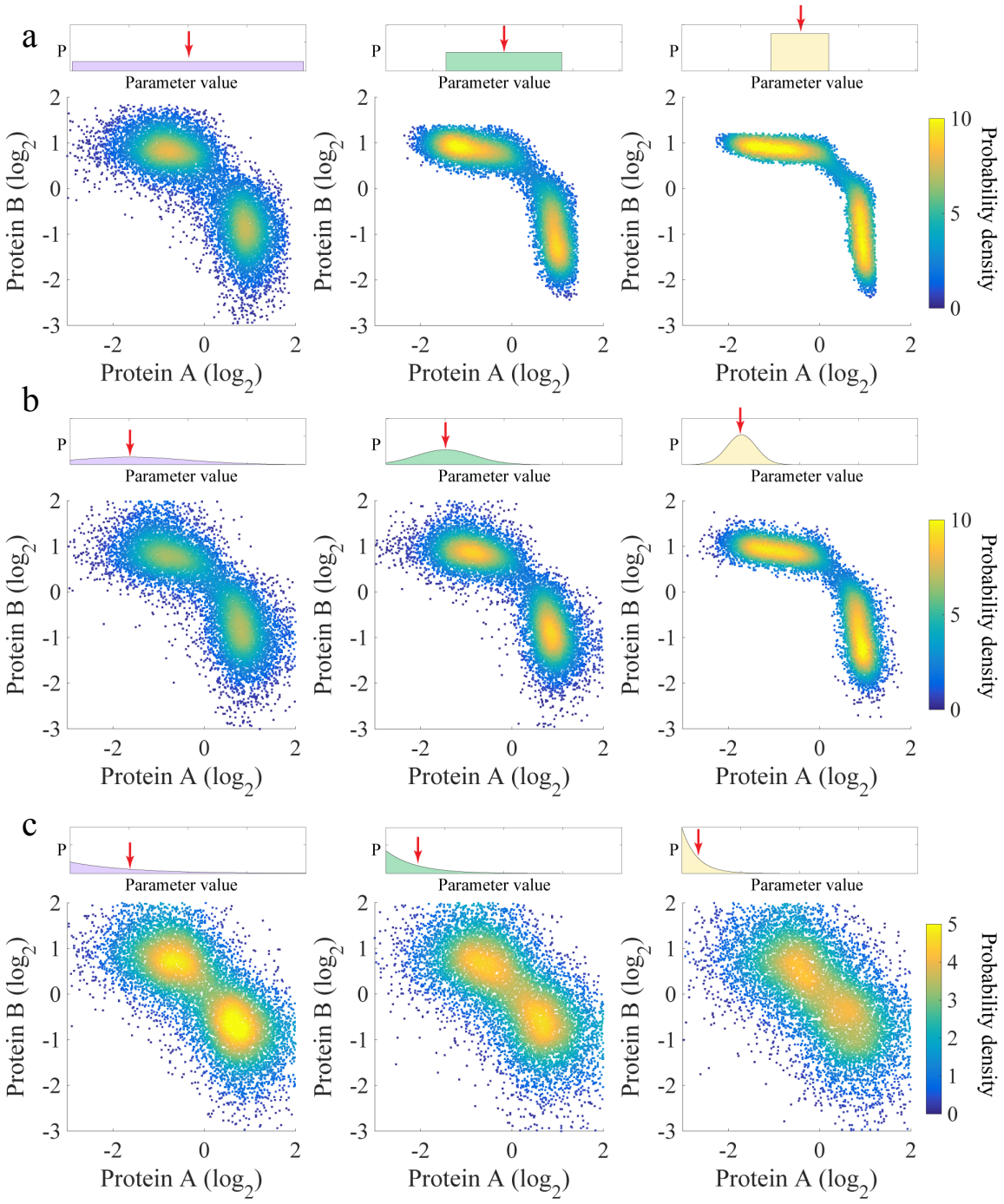


Figure 3.4 The gene states of the toggle-switch motif are robust against different types of distributions used to sample the parameters.

(a) Uniform distributions in three different ranges were used to sample the kinetic parameters of the RACIPE models. The top panels show the range of the distribution

(left panel: the full range; middle panel: half; right panel: 1/4). The bottom panels show the probability density maps of the gene expression data from all the RACIPE models. Similarly, panels **(b)** and **(c)** show the use of a Gaussian distribution and an exponential distribution, respectively. For the Gaussian distribution (b), its standard deviation was shrunk by a factor of two from left to right. For the Exponential distribution (c), its mean was reduced by a factor of two from left to right.

To evaluate the effectiveness of RACIPE on larger circuits, we further applied the method to circuits with two to five coupled toggle-switch (CTS) motifs (Figure 3.5). Different from the above simple circuit motifs, the gene expression data obtained by RACIPE for these CTS motifs are now high-dimensional; thus in the scatter plot analysis we projected these data onto the first two principal components by principal component analysis (PCA). For each circuit, we observed distinct gene states from PCA for the RACIPE models (Figure 3.5a). More interestingly, the number of gene states found via PCA increases by one each time one more toggle switch is added to the circuit. Moreover, we applied HCA to the gene expression data, from which we identified the same gene states as from PCA (Figure 3.5b). At this stage, we can also assign high (red circles), intermediate (blue circles) or low expression (black circles) to each gene for every gene state. Unlike in Boolean network models, the assignment in RACIPE is based on the distribution of expression pattern from all the models in the ensemble (Figure S2.6 and S2.7).

We can easily understand the meaning of each gene state. In each case, the rightmost cluster in the scatter plot (Figure 3.5a) corresponds to the topmost cluster in the heatmap (Figure 3.5b), a state where all the A genes have high expression and all the B

genes have low expression. Similarly, the leftmost cluster in the scatter plot corresponds to the bottommost cluster in the heatmap. These two clusters are the most probable ones, and represent the two extreme states of the coupled toggle switch network. As also illustrated in the scatter plots, for circuits with additional toggle switches, these two states separate from each other and the circuit now allows intermediate states. By closely examining these intermediate states, we found that they (from top to bottom) correspond to a cascade of flips of the state of each consecutive toggle switch. This explains why we observe one additional gene state every time we include an additional toggle-switch motif. In addition, intermediate expression levels were frequently observed for genes lying in the middle toggle-switch motifs, instead of those at the edge. The tests on CTS circuits demonstrate again the power of RACIPE in identifying robust features of a complex circuit.

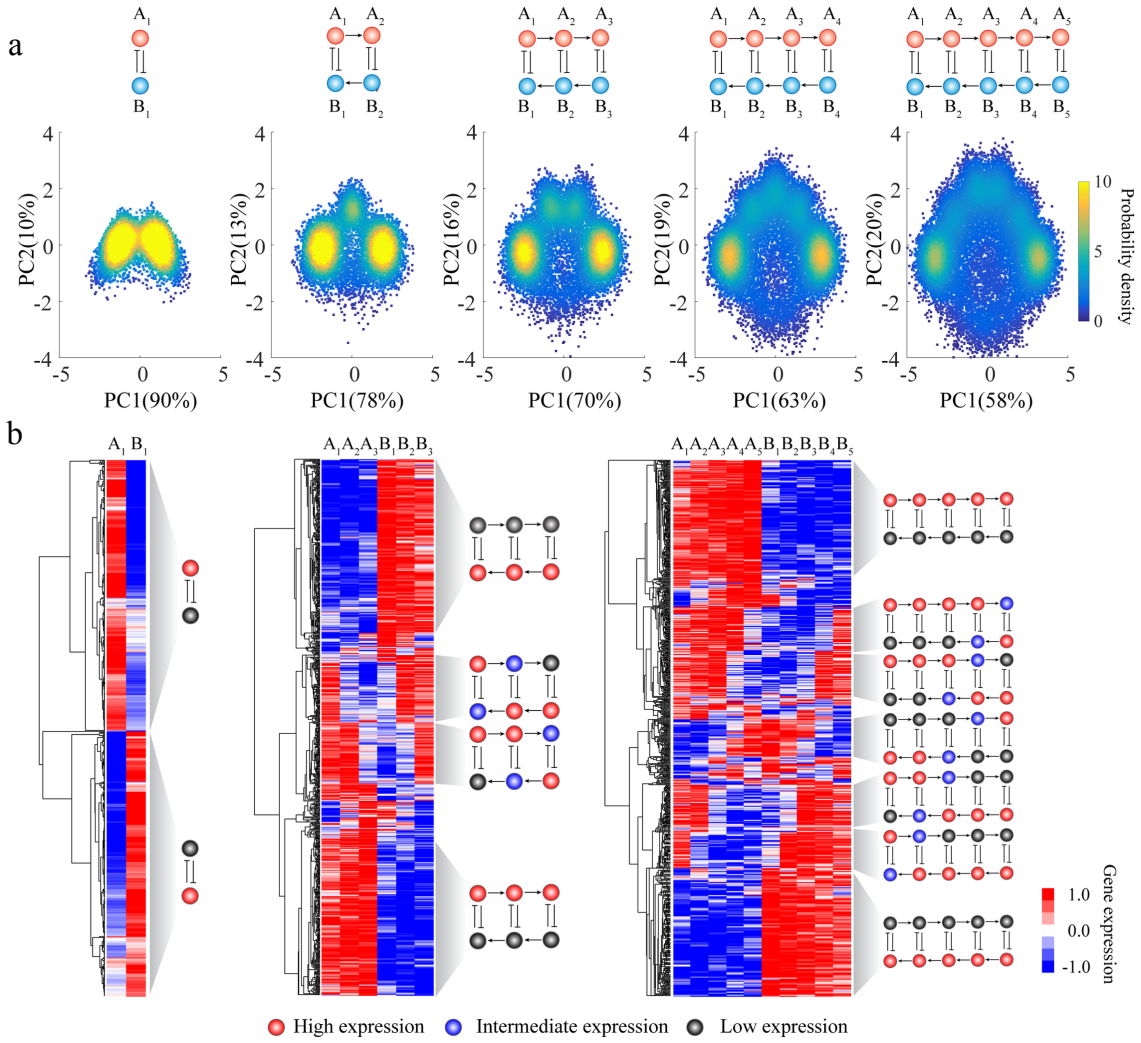


Figure 3.5 Application of RACIPE to coupled toggle-switch circuits.

RACIPE was tested on coupled toggle-switch circuits, as illustrated at the top of the figure. **(a)** 2D probability density map of the RACIPE-predicted gene expression data projected to the 1st and 2nd principal component axes. **(b)** Average linkage hierarchical clustering analysis of the gene expression data from all the RACIPE models using the Euclidean distance. Each column corresponds to a gene, while each row corresponds to a stable steady state. The clustering analysis allows the identification of several robust gene states, whose characteristics were illustrated as circuit cartoons to the right of the heatmap. The expression levels of each gene in these gene states are illustrated as low (grey), intermediate (blue), or high (red, see Figure S2.6 and S2.7 for the details).

3.3.2. Epithelial-to-Mesenchymal circuit

3.3.2.1. Introduction

EMT, as mentioned above, is crucial for embryonic development, wound healing and metastasis[155], a major cause for 90% cancer-related deaths[26]. Cells can undergo either a complete EMT to acquire mesenchymal phenotype or partial EMT to attain hybrid E/M phenotype[28,156], which maintains both E and M traits. Transitions among the Epithelial (E), Mesenchymal (M) and hybrid epithelial/mesenchymal (E/M) phenotypes have been widely studied either experimentally or theoretically[28]. Here, we utilized data from the literature and Ingenuity Pathway Analysis to construct a larger core gene regulatory circuit model of EMT (Figure 3.6a), which contains 13 transcriptional factors (TFs), 9 microRNAs (miRs) and 82 regulatory links among them. Among the gene components, we have two biomarkers – CDH1 and VIM – that are commonly used to distinguish different phenotypes during EMT, and one signaling gene TGF-β. The circuit is a much-extended version of several previous EMT models[12,40], which consist of only four gene families and one input signal. It is similar in terms of scale to a recently proposed Boolean model of EMT[157], but as stressed here our models allow for continuous expression levels.

3.3.2.2. Methods

Construction of EMT circuit

In our previous studies[28,40], we constructed a coarse-grained core gene regulatory circuit of EMT, consisting of two transcription factor (TF) families (SNAIL

and ZEB) and two microRNA (miR) families (miR-34 and miR-200) (Figure 2.2). In addition, we also included an input node, representing the integration of multiple cell signaling pathways, such as HGF, NF-κB, WNT, TGF-β and HIF1-α, into SNAIL. The outputs of the circuit are two commonly used biomarkers of EMT – CDH1 and VIM.

Here, we expanded the size of the EMT circuit on the basis of the core EMT module as follows. For each gene family, we considered individual members as separate nodes: SNAI1, SNAI2, TWIST1, TWIST2 for the SNAIL family, ZEB1 and ZEB2 for the ZEB family, miR-200a, miR-200b, miR-200c, miR-141 and miR-205 for the miR-200 family, and miR-34a for the miR-34 family. The input and output nodes, TGF-beta, CDH1 and VIM, were explicitly included in the new circuit model. We also included additional genes (FOXC2, KLF8, miR-101, miR-30c, miR-9, TCF3 and GSC) and the directed interactions among all the gene components according to recent experimental evidences[158,159], theoretical studies[13,160,161] and IPA(IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity).

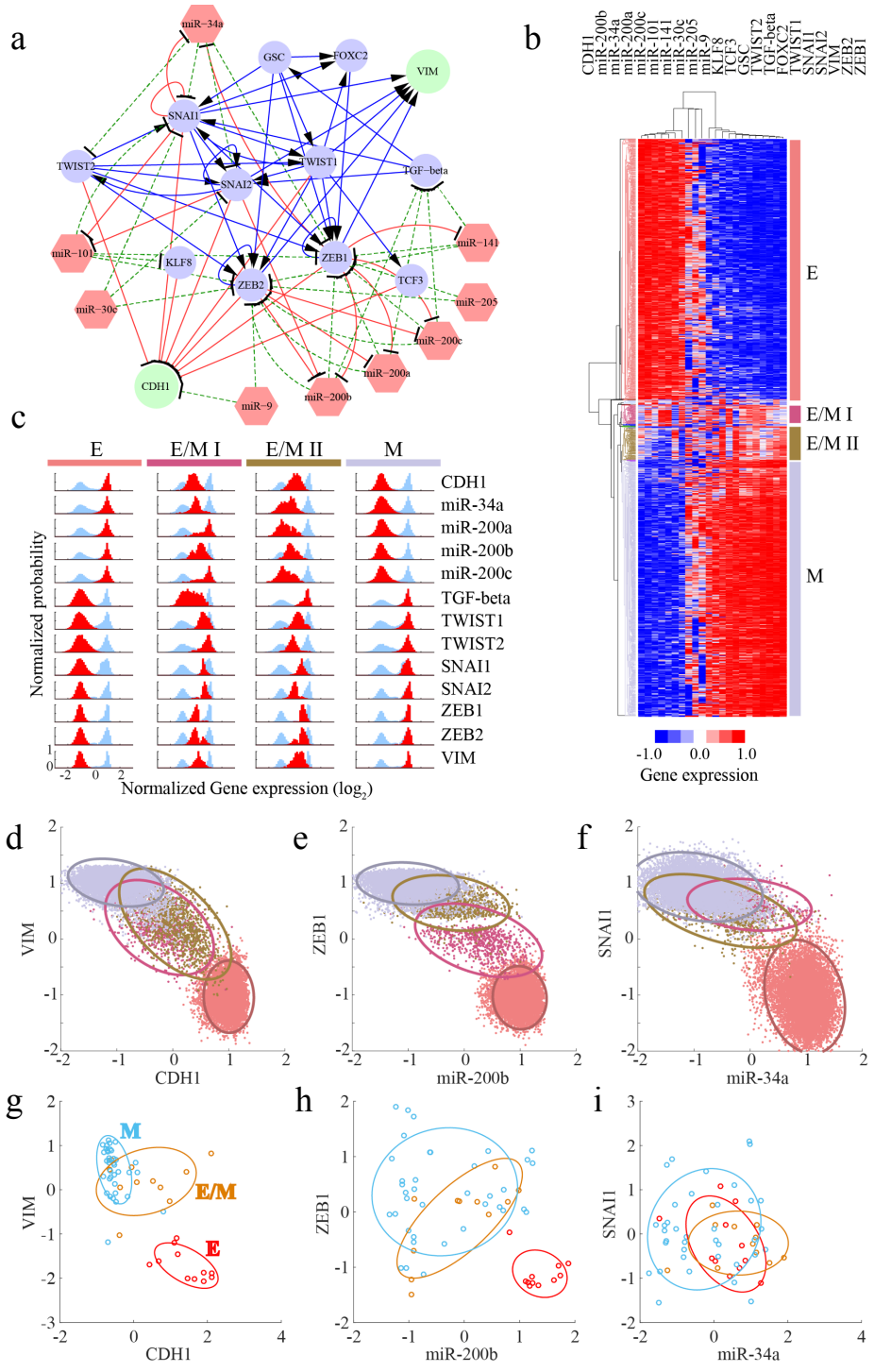


Figure 3.6 RAICPE identifies multiple EMT cell states from gene network analysis.

(a) A proposed Epithelial-to-Mesenchymal Transition (EMT) circuit is constructed according to the literature; the circuit consists of 13 transcriptional factors (circles), 9 microRNAs (red hexagons) and 82 regulatory links. The blue solid lines and arrows represent transcriptional activations the red solid lines and bars represent transcriptional inhibition, and the green dashed lines and bars stand for translational inhibition. Two readout genes CDH1 and VIM are shown as green circles while the other transcriptional factors are shown in blue. **(b)** Average linkage hierarchical clustering analysis of the gene expression data from all the RACIPE models using the Euclidean distance. Each column corresponds to a gene, and each row corresponds to a stable steady state. Four major gene states were identified and highlighted by different colors. According to the expression levels of CDH1 and VIM, the four gene states were associated with epithelial (E in red), mesenchymal (M in grey) and two hybrid epithelial/mesenchymal (E/M I in purple and E/M II in brown) phenotypes. **(c)** The gene expression distribution of each gene state. The gene expression distribution of each gene for all of the RACIPE models is shown in blue, while that for each gene state is shown in red (50 bins are used to calculate the histogram of each distribution). For clarity, each distribution is normalized by its maximum probability. Each row represents a gene and each column represents a gene state. **(d-f)** Gene expression data were projected to either CDH1/VIM, miR-200b/ZEB1, or miR-34a/SNAI1 axes. Different gene states are highlighted by the corresponding colors and enclosed by the ellipses. **(g-i)** Transcriptomics data from the NCI-60 cell lines were projected to either CDH1/VIM, miR-200b/ZEB1, or miR-34a/SNAI1 axes. The NCI-60 cell lines have been grouped into E, E/M and M phenotypes according to the ratio of the protein levels of CDH1 and VIM. Different gene states are highlighted by the corresponding colors and enclosed by the ellipses.

Mathematical Model of EMT circuit

For simplicity, we modeled the EMT circuit with the same approach as above, i.e. all the genetic components were coupled with Hill functions, typical of transcriptional control. This may not be completely accurate for translational regulation by microRNA (miR), but we leave this complication for future study.

$$
\begin{aligned}
\dot{X_1} &= G_{X_1}(H^S(X_7, X_7X_{10}, n_{X_7X_1}, \lambda^+_{X_7X_1})/\lambda^+_{X_7X_1})(H^S(X_{20}, X_{20}X_{10}, n_{X_{20}X_1}, \lambda^+_{X_{20}X_1})/\lambda^+_{X_{20}X_1})(H^S(X_{22}, X_{22}X_{10}, n_{X_{22}X_1}, \lambda^+_{X_{22}X_1})/\lambda^+_{X_{22}X_1}) - k_{X_1}X_1\\
\dot{X_2} &= G_{X_2}(H^S(X_1, X_1X_{20}, n_{X_1X_2}, \lambda^+_{X_1X_2})/\lambda^+_{X_1X_2})H^S(X_5, X_5X_{20}, n_{X_5X_2}, \lambda^-_{X_5X_2})H^S(X_8, X_8X_{20}, n_{X_8X_2}, \lambda^-_{X_8X_2})H^S(X_{10}, X_{10}X_{20}, n_{X_{10}X_2}, \lambda^-_{X_{10}X_2})\\
&\quad H^S(X_{11}, X_{11}X_{20}, n_{X_{11}X_2}, \lambda^-_{X_{11}X_2})H^S(X_{12}, X_{12}X_{20}, n_{X_{12}X_2}, \lambda^-_{X_{12}X_2})H^S(X_{13}, X_{13}X_{20}, n_{X_{13}X_2}, \lambda^-_{X_{13}X_2})H^S(X_{16}, X_{16}X_{20}, n_{X_{16}X_2}, \lambda^-_{X_{16}X_2})\\
&\quad (H^S(X_7, X_7X_{20}, n_{X_7X_2}, \lambda^+_{X_7X_2})/\lambda^+_{X_7X_2})(H^S(X_{20}, X_{20}X_{20}, n_{X_{20}X_2}, \lambda^+_{X_{20}X_2})/\lambda^+_{X_{20}X_2})(H^S(X_{17}, X_{17}X_{20}, n_{X_{17}X_2}, \lambda^+_{X_{17}X_2})/\lambda^+_{X_{17}X_2})\\
&\quad (H^S(X_{22}, X_{22}X_{20}, n_{X_{22}X_2}, \lambda^+_{X_{22}X_2})/\lambda^+_{X_{22}X_2}) - k_{X_2}X_2\\
\dot{X_3} &= G_{X_3}H^S(X_5, X_5X_{30}, n_{X_5X_3}, \lambda^-_{X_5X_3}) - k_{X_3}X_3\\
\dot{X_4} &= G_{X_4}H^S(X_3, X_3X_{40}, n_{X_3X_4}, \lambda^-_{X_3X_4})H^S(X_{18}, X_{18}X_{40}, n_{X_{18}X_4}, \lambda^-_{X_{18}X_4})H^S(X_7, X_7X_{40}, n_{X_7X_4}, \lambda^-_{X_7X_4})H^S(X_{15}, X_{15}X_{40}, n_{X_{15}X_4}, \lambda^-_{X_{15}X_4})\\
&\quad H^S(X_{21}, X_{21}X_{40}, n_{X_{21}X_4}, \lambda^-_{X_{21}X_4})H^S(X_{20}, X_{20}X_{40}, n_{X_{20}X_4}, \lambda^-_{X_{20}X_4})H^S(X_{17}, X_{17}X_{40}, n_{X_{17}X_4}, \lambda^-_{X_{17}X_4})H^S(X_2, X_2X_{40}, n_{X_2X_4}, \lambda^-_{X_2X_4})\\
&\quad H^S(X_6, X_6X_{40}, n_{X_6X_4}, \lambda^-_{X_6X_4}) - k_{X_4}X_4\\
\dot{X_5} &= G_{X_5}H^S(X_7, X_7X_{50}, n_{X_7X_5}, \lambda^-_{X_7X_5})H^S(X_{15}, X_{15}X_{50}, n_{X_{15}X_5}, \lambda^-_{X_{15}X_5}) - k_{X_5}X_5\\
\dot{X_6} &= G_{X_6}H^S(X_5, X_5X_{60}, n_{X_5X_6}, \lambda^-_{X_5X_6})H^S(X_8, X_8X_{60}, n_{X_8X_6}, \lambda^-_{X_8X_6})H^S(X_{10}, X_{10}X_{60}, n_{X_{10}X_6}, \lambda^-_{X_{10}X_6})H^S(X_{11}, X_{11}X_{60}, n_{X_{11}X_6}, \lambda^-_{X_{11}X_6})\\
&\quad H^S(X_{12}, X_{12}X_{60}, n_{X_{12}X_6}, \lambda^-_{X_{12}X_6})H^S(X_{13}, X_{13}X_{60}, n_{X_{13}X_6}, \lambda^-_{X_{13}X_6})H^S(X_{14}, X_{14}X_{60}, n_{X_{14}X_6}, \lambda^-_{X_{14}X_6})H^S(X_{18}, X_{18}X_{60}, n_{X_{18}X_6}, \lambda^-_{X_{18}X_6})\\
&\quad (H^S(X_7, X_7X_{60}, n_{X_7X_6}, \lambda^+_{X_7X_6})/\lambda^+_{X_7X_6})(H^S(X_{20}, X_{20}X_{60}, n_{X_{20}X_6}, \lambda^+_{X_{20}X_6})/\lambda^+_{X_{20}X_6})(H^S(X_{17}, X_{17}X_{60}, n_{X_{17}X_6}, \lambda^+_{X_{17}X_6})/\lambda^+_{X_{17}X_6})\\
&\quad (H^S(X_{22}, X_{22}X_{60}, n_{X_{22}X_6}, \lambda^+_{X_{22}X_6})/\lambda^+_{X_{22}X_6}) - k_{X_6}X_6\\
\dot{X_7} &= G_{X_7}H^S(X_5, X_5X_{70}, n_{X_5X_7}, \lambda^-_{X_5X_7})H^S(X_{14}, X_{14}X_{70}, n_{X_{14}X_7}, \lambda^-_{X_{14}X_7})H^S(X_{16}, X_{16}X_{70}, n_{X_{16}X_7}, \lambda^-_{X_{16}X_7})H^S(X_7, X_7X_{70}, n_{X_7X_7}, \lambda_{X_7X_7}\\
&\quad (H^S(X_{20}, X_{20}X_{70}, n_{X_{20}X_7}, \lambda^+_{X_{20}X_7})/\lambda^+_{X_{20}X_7})(H^S(X_{17}, X_{17}X_{70}, n_{X_{17}X_7}, \lambda^+_{X_{17}X_7})/\lambda^+_{X_{17}X_7})(H^S(X_{22}, X_{22}X_{70}, n_{X_{22}X_7}, \lambda^+_{X_{22}X_7})/\lambda^+_{X_{22}X_7})\\
&\quad (H^S(X_9, X_9X_{70}, n_{X_9X_7}, \lambda^+_{X_9X_7})/\lambda^+_{X_9X_7}) - k_{X_7}X_7\\
\dot{X_8} &= G_{X_8}H^S(X_2, X_2X_{80}, n_{X_2X_8}, \lambda^-_{X_2X_8}) - k_{X_8}X_8\\
\dot{X_9} &= G_{X_9}H^S(X_8, X_8X_{90}, n_{X_8X_9}, \lambda^-_{X_8X_9})H^S(X_{10}, X_{10}X_{90}, n_{X_{10}X_9}, \lambda^-_{X_{10}X_9})H^S(X_{11}, X_{11}X_{90}, n_{X_{11}X_9}, \lambda^-_{X_{11}X_9})H^S(X_{12}, X_{12}X_{90}, n_{X_{12}X_9}, \lambda^-_{X_{12}X_9})\\
&\quad - k_{X_9}X_9\\
\dot{X_{10}} &= G_{X_{10}}H^S(X_2, X_2X_{100}, n_{X_2X_{10}}, \lambda^-_{X_2X_{10}})H^S(X_6, X_6X_{100}, n_{X_6X_{10}}, \lambda^-_{X_6X_{10}}) - k_{X_{10}}X_{10}\\
\dot{X_{11}} &= G_{X_{11}}H^S(X_{15}, X_{15}X_{110}, n_{X_{15}X_{11}}, \lambda^-_{X_{15}X_{11}})H^S(X_2, X_2X_{110}, n_{X_2X_{11}}, \lambda^-_{X_2X_{11}})H^S(X_6, X_6X_{110}, n_{X_6X_{11}}, \lambda^-_{X_6X_{11}}) - k_{X_{11}}X_{11}\\
\dot{X_{12}} &= G_{X_{12}}H^S(X_2, X_2X_{120}, n_{X_2X_{12}}, \lambda^-_{X_2X_{12}})H^S(X_6, X_6X_{120}, n_{X_6X_{12}}, \lambda^-_{X_6X_{12}}) - k_{X_{12}}X_{12}\\
\dot{X_{13}} &= G_{X_{13}} - k_{X_{13}}X_{13}\\
\dot{X_{14}} &= G_{X_{14}} - k_{X_{14}}X_{14}\\
\dot{X_{15}} &= G_{X_{15}}H^S(X_{14}, X_{14}X_{150}, n_{X_{14}X_{15}}, \lambda^-_{X_{14}X_{15}})H^S(X_{16}, X_{16}X_{150}, n_{X_{16}X_{15}}, \lambda^-_{X_{16}X_{15}})(H^S(X_7, X_7X_{150}, n_{X_7X_{15}}, \lambda^+_{X_7X_{15}})/\lambda^+_{X_7X_{15}})\\
&\quad (H^S(X_{15}, X_{15}X_{150}, n_{X_{15}X_{15}}, \lambda^+_{X_{15}X_{15}})/\lambda^+_{X_{15}X_{15}})(H^S(X_{20}, X_{20}X_{150}, n_{X_{20}X_{15}}, \lambda^+_{X_{20}X_{15}})/\lambda^+_{X_{20}X_{15}})(H^S(X_{17}, X_{17}X_{150}, n_{X_{17}X_{15}}, \lambda^+_{X_{17}X_{15}})/\lambda^+_{X_{17}X_{15}})\\
&\quad (H^S(X_9, X_9X_{150}, n_{X_9X_{15}}, \lambda^+_{X_9X_{15}})/\lambda^+_{X_9X_{15}}) - k_{X_{15}}X_{15}\\
\dot{X_{16}} &= G_{X_{16}}H^S(X_7, X_7X_{160}, n_{X_7X_{16}}, \lambda^-_{X_7X_{16}})H^S(X_2, X_2X_{160}, n_{X_2X_{16}}, \lambda^-_{X_2X_{16}}) - k_{X_{16}}X_{16}\\
\dot{X_{17}} &= G_{X_{17}}H^S(X_{16}, X_{16}X_{170}, n_{X_{16}X_{17}}, \lambda^-_{X_{16}X_{17}})(H^S(X_{15}, X_{15}X_{170}, n_{X_{15}X_{17}}, \lambda^+_{X_{15}X_{17}})/\lambda^+_{X_{15}X_{17}}) - k_{X_{17}}X_{17}\\
\dot{X_{18}} &= G_{X_{18}} - k_{X_{18}}X_{18}\\
\dot{X_{19}} &= G_{X_{19}}(H^S(X_7, X_7X_{190}, n_{X_7X_{19}}, \lambda^+_{X_7X_{19}})/\lambda^+_{X_7X_{19}})(H^S(X_{15}, X_{15}X_{190}, n_{X_{15}X_{19}}, \lambda^+_{X_{15}X_{19}})/\lambda^+_{X_{15}X_{19}})(H^S(X_{20}, X_{20}X_{190}, n_{X_{20}X_{19}}, \lambda^+_{X_{20}X_{19}})/\lambda^+_{X_{20}X_{19}})\\
&\quad (H^S(X_2, X_2X_{190}, n_{X_2X_{19}}, \lambda^+_{X_2X_{19}})/\lambda^+_{X_2X_{19}})\\
&\quad (H^S(X_6, X_6X_{190}, n_{X_6X_{19}}, \lambda^+_{X_6X_{19}})/\lambda^+_{X_6X_{19}}) - k_{X_{19}}X_{19}\\
\dot{X_{20}} &= G_{X_{20}}(H^S(X_7, X_7X_{200}, n_{X_7X_{20}}, \lambda^+_{X_7X_{20}})/\lambda^+_{X_7X_{20}})(H^S(X_{17}, X_{17}X_{200}, n_{X_{17}X_{20}}, \lambda^+_{X_{17}X_{20}})/\lambda^+_{X_{17}X_{20}})(H^S(X_{22}, X_{22}X_{200}, n_{X_{22}X_{20}}, \lambda^+_{X_{22}X_{20}})/\lambda^+_{X_{22}X_{20}})\\
&\quad - k_{X_{20}}X_{20}\\
\dot{X_{21}} &= G_{X_{21}}(H^S(X_{20}, X_{20}X_{210}, n_{X_{20}X_{21}}, \lambda^+_{X_{20}X_{21}})/\lambda^+_{X_{20}X_{21}}) - k_{X_{21}}X_{21}\\
\dot{X_{22}} &= G_{X_{22}}(H^S(X_9, X_9X_{220}, n_{X_9X_{22}}, \lambda^+_{X_9X_{22}})/\lambda^+_{X_9X_{22}}) - k_{X_{22}}X_{22}
\end{aligned}
\quad , (26)
$$

where G_i represents the maximum production rate of gene i, and k_i represents the degradation rate of gene i. For the regulatory link (either activation or inhibition) from gene i to gene j, ij_0 represents the threshold level of gene i, n_{ij} is the Hill coefficient, and

λ_{ij} is the maximum fold change of the gene j level caused by gene i. $\lambda_{ij} < 1$ for an inhibitory regulation, referred to as λ_{ij}^{-}, while $\lambda_{ij} > 1$ for an excitatory regulation, referred to as λ_{ij}^{+}. For the model of the EMT circuit, X_i (i = 1, 2 … 22) stand for the expression levels of FOXC2, ZEB1, KLF8, CDH1, miR-101, ZEB2, SNAI1, miR-141, TGF-beta, miR-200a, miR-200b, miR-200c, miR-205, miR-30c, SNAI2, miR-34a, TWIST2, miR-9, VIM, TWIST1, TCF3, GSC, respectively.

Analysis of EMT gene expression data from NCI-60 cell lines

The gene transcript z scores of CDH1, VIM, ZEB1, miR-200b, SNAIL, miR-34a were downloaded from Cellminer[162] and categorized into epithelial, hybrid E/M and mesenchymal sets based on the ratio of the E-cadherin and Vimentin levels[93].

3.3.2.3. Results

Even with this simplification, RACIPE can already provide insightful information of the EMT regulation. Consistent with what we learned from the test cases, unsupervised HCA of the RACIPE gene expression data can reveals distinct gene states (Figure 3.6b). Here there are four such states. We can map these gene states to different cell phenotypes possible during EMT – an E phenotype with high expression of the miRs, low expression of TFs, and $CDH1^{HI}VIM^{LO}$; a M phenotype with low expression of the miRs, high expressions of TFs, and $CDH1^{LO}VIM^{HI}$; and two hybrid E/M phenotypes with intermediate expression of both miRs, TFs and CDH1/VIM. The E/M I state lies closer to the E state, and the E/M II state lies closer to the M state. More intriguingly, we found

SNAI1 and SNAI2 become highly expressed in the E/M I phenotype while ZEB1 and ZEB2 are not fully expressed until the E/M II or the M phenotype (Figure 3.6c), which is a possibility supported by recent experimental results[12].

Moreover, RACIPE can help to find genes of similar function and filter out less important genes in the core circuit. As shown in Figure 3.6b, genes are grouped into two major clusters according to their expression levels throughout all the RACIPE models – miRs/CDH1 and TFs/VIM. The former genes are highly expressed mainly in E phenotypes while the latter are highly expressed in M phenotypes. We also found three microRNAs (miR-30c, miR-205 and miR-9) to be randomly expressed in the RACIPE models, indicating these three microRNAs are less important to these EMT phenotypes. From the topology of the circuit, we see that these three microRNAs lack feedback regulation and act solely as inputs.

A typical approach taken in cell biology is to use two biomarkers to identify cells of different states in a mixed population by fluorescence-activated cell sorting (FACS). To mimic the analysis, we projected the gene expression data of the RACIPE models onto the two axes of important genes, as shown in the scatter plots in Figure 3.6d-f. In all of the three cases, the E and the M phenotypes can be distinguished. However, for the hybrid phenotypes, the E/M I and the E/M II states overlap in the CDH1-VIM plot (Figure 3.6d). These two hybrid phenotypes can be separated more easily in the ZEB1-miR200b plot (Figure 3.6e). In the SNAIL1-miR34a plot (Figure 3.6f), however, the two E/M states overlap with the M state. The theoretical prediction that the SNAIl1-miR34a axis is less efficient at distinguishing the states is supported by transcriptomics data from

the NCI-60 cell lines[162] (Figure 3.6g-i). We see here that either VIM-CDH1 or the ZEB1-miR200b axes are indeed better than the SNAIL1-miR34a axes in separating different EMT phenotypes. Our results are also consistent with our previous theoretical finding that ZEB1 is more crucial than SNAIL1 in the decision-making of EMT[12].

3.4. Discussion

Recently, the rapid development of genomic profiling tools has allowed the mapping of gene regulatory networks. Yet, it remains a challenge to understand the operating mechanisms and the design principles of these networks. Conventional computational modeling methods provide insightful information; however, their prediction power is usually limited by the incompleteness of the network structure and the absence of reliable kinetics. To deal with these issues, we have developed a new computational modeling method, called RACIPE, which allows unbiased predictions of the dynamic behaviors of a complex gene regulatory circuit. Compared to traditional methods, RACIPE uniquely generates an ensemble of models with distinct kinetic parameters. These models can faithfully represent the circuit topology and meanwhile capture the heterogeneity in the kinetics of the genetic regulation. By modeling the dynamics of every RACIPE model, we can utilize statistical analysis tools to identify the robust features of network dynamics. We have successfully tested RACIPE on several theoretical circuit motifs and a proposed core Epithelial-to-Mesenchymal Transition (EMT) gene regulatory circuit. In each circuit, RACIPE is capable of predicting the

relevant gene states and providing insights into the regulatory mechanism of the decision-making among gene states.

Unlike other methods that utilize randomization strategies to explore the parameter sensitivity for gene circuit[141–144], RACIPE adopts a more carefully designed sampling strategy to randomize circuit parameters over a wide range, but meanwhile to satisfy the half-functional rule to gain a comprehensive understanding of circuit dynamics. Instead of looking for the sensitivity of the circuit function to parameter variations [141,144] and the parameters best fitting the experimental data[142,143], we focused on uncovering conserved features from the ensemble of RACIPE models. This was carried out by standard statistical learning methods such as hierarchical clustering analysis. We showed the power of RACIPE to predict the robust gene states for a circuit with a given topology. Also, conceptually similar to the mixed-effects models used to describe a cell population for a very simple system [143], i.e. a one-gene transcription without a regulator, RACIPE could be potentially applied to a very large gene circuit to describe the gene expression dynamics of a cell population with an ensemble of models - an aspect we will work on in our future study. Moreover, it is easy to implement gene modifications such as knockdown or overexpression treatments with the RACIPE method to learn the significance of each gene or link in the circuit. Therefore, RACIPE provides a new way to model a gene circuit without knowing the detailed circuit parameters.

Another parameter-independent approach people often use for gene circuit modeling is Boolean network model[163], which digitalizes the gene expression into on and off states and uses logic functions to describe the combinatorial effects of regulators

to their targets. Compared with the Boolean network model, RACIPE is a continuous method, so it is not restricted to the on and off values. Instead, RACIPE enables us to find the intermediate levels of gene expressions beyond the on and off states, as we showed in Figure 3.5 and Figure 3.6. From the ensemble of RACIPE models, we can predict the expression distribution of each gene, which can be directly compared with experimental expression data. The comparison will allow us to further refine the core circuit. In addition, in RACIPE, we not only obtain *in silico* gene expression data, but we also know the kinetic parameters for each model. From these parameter data, we can directly compare the parameter distributions for different gene states, from which we can learn the driving parameters that are responsible for the transitions among the states.

To conclude, here we have introduced a new theoretical modeling method, RACIPE, to unbiasedly study the behavior of a core gene regulatory circuit under the presence of intrinsic or extrinsic fluctuations. These fluctuations could represent different signaling environments, epigenetic states, and/or genetic backgrounds of the core circuit and can cause cell-cell heterogeneity in a population. By approximating these fluctuations as variations of the model parameters, RACIPE provides a straightforward way to understand the heterogeneity and to explain further how gene circuits can perform robust functions under such conditions. Moreover, RACIPE uniquely generates a large set *in silico* expression data, which can be directly compared with experimental data using common bioinformatics tools. RACIPE enables the connection of traditional circuit-based bottom-up approach with profiling-based top-down approach. We expect RACIPE to be a powerful method to identify the role of network topology in determining network operating principles.

Chapter 4

Decoding Stem Cell Regulatory Circuit by Random Circuit Perturbation Method

Stem cells are capable to perform precise and robust fate decisions during development; however, how the underlying gene regulatory network reliably specify the fates under the influence of intrinsic and extrinsic noises is not well understood. Here, we applied Random Circuit Perturbation (RACIPE) method on a proposed stem cell gene regulatory circuit, and found that the stable states of all models converge to experimentally observed gene state clusters even when the parameters are strongly perturbed. Compared to circuits with random topology, the stem cell circuit allows for more robust gene states and is less likely to have oscillatory/chaotic dynamics. From parametric perturbations for all the RACIPE models, we identified key genes and the concomitant hierarchical structure of stem cell differentiation. Our results suggest that dynamics of the stem cell circuit is mainly maintained by its well-evolved topology, instead by detailed circuit parameters.

4.1. Introduction

The differentiation of stem cells is an intense research area in developmental biology[164] and regenerative medicine[165]. Stem cells are unique in that they are capable of generating all types of cells in a precise and organized manner[166], and the dysregulation of the differentiation results in early fetal deaths or severe diseases. Thus, the understanding of the mechanism of stem cell differentiation is crucial to the reprogramming of differentiated cells[167] and the stem cell-based therapies[168]. Due to their essential role in survival for all multicellular organisms, the mechanism of the differentiation is presumably highly conserved and able to precisely make decisions for each step lineage specification. However, recent evidences suggest that some development-related regulators, such as Nanog, are expressed in large variation[169–173]. Moreover, stem cells have been shown to have heterogeneous gene expression in culture[170–172], and cells with different expressions can sometimes convert into each other[171,172]. The observed heterogeneity seems to put precise fate decision at risk. This paradox may be reconciled by two possible explanations - either the regulatory mechanism of stem cell differentiation is robust in performing its functions against gene noises[174], or it takes advantage of the noises for its own functions[175]. A better understanding to this question, and thereafter to the role of gene noise in decision-making, will be especially helpful for decoding the design principle of the underlying gene regulatory network for stem cell differentiation.

Computational methods are well utilized in this field to study the regulatory mechanism of stem cell differentiation on the level of gene regulatory network, which

now is able to be constructed or inferred by current molecular profiling techniques[176,177]. Considering the complexity of these network, the bottom-up modeling approaches of the core gene regulatory circuit with master regulators such as Oct4, Cdx2 etc. are wildly studied by either deterministic models[178,179] or stochastic models[180–182]. Also, the dynamic Bayesian network (DBN) method[183] and the landscape and kinetic path approach[184] have been used to explore larger gene networks of the stem cell. However, it still remains a challenge to fully understand the regulatory mechanisms by current quantitative methods, due to the lack of the experimental kinetic data for most of the regulatory processes, and also the potential over-simplified standalone circuits without considering the effects of many other genes and the complex, variable cellular microenvironment.

In our previous work[185], we developed a new computational method, named random circuit perturbation (RACIPE), for modeling the possible dynamic behavior defined by the topology of a gene regulatory circuit without the prior knowledge of the accurate parameters of the circuit. RACIPE generates an ensemble of mathematical models (Figure 3.1), each of which has a different set of kinetic parameters representing intrinsic and extrinsic noises caused by variations of signaling states, epigenetic states, and genetic backgrounds (including cells with genetic mutations leading to disease). The parameters of each model differ by one or two orders of magnitude and are generated under a specially designed parameter sampling scheme to capture the key role of circuit topology. The computationally generated ensemble predicts the expressions of the genes in the circuit under variable environments, which is subject to statistical methods to identify the most probable and conservative features. More importantly, RACIPE

provides a physical model underlying the gene expression data, enable us to decode the design principle and to evaluate the robustness of a core gene circuit.

Here, we applied RACIPE on the core regulatory circuit of stem cell differentiation, and identified robust gene expression patterns in consistent with single cell gene expression data. We also demonstrated that the topology of stem cell circuit is well evolved to perform its functions. Furthermore, *in silico* perturbation treatments of the ensemble of RACIPE models reveal the hidden hierarchical structure in the circuit, allowing a two-step differentiation process.

4.2. Methods

Application of RACIPE on stem cell circuit

Based on previous studies[183,184,186,187], we constructed a core gene regulatory circuit for mouse stem cell differentiation, composed of eight transcription factors (*Oct4*, *Cdx2*, etc.), a protein complex (OCT4-SOX2) and 25 regulations among them (Figure 4.1a). Due to the complexity of the regulatory links among genes, the operative functions of this circuit have to date remained elusive. The dynamic behavior of stem cell circuit can be described by the deterministic rate equations as follows:

$$
\begin{aligned}
\dot{X}_1 &= G_{X_1} H^S(X_1, X_1X_{10}, n_{X_1X_1}, \lambda^+_{X_1X_1}) H^S(X_5, X_5X_{10}, n_{X_5X_1}, \lambda^-_{X_5X_1}) H^S(X_7, X_7X_{10}, n_{X_7X_1}, \lambda^+_{X_7X_1}) H^S(X_8, X_8X_{10}, n_{X_8X_1}, \lambda^-_{X_8X_1}) \\
&\quad / (\lambda^+_{X_7X_1}\lambda^+_{X_7X_1}) - k_{X_1}X_1 \\
\dot{X}_2 &= G_{X_2} H^S(X_1, X_1X_{20}, n_{X_1X_2}, \lambda^+_{X_1X_2}) H^S(X_3, X_3X_{20}, n_{X_3X_2}, \lambda^+_{X_3X_2}) / (\lambda^+_{X_1X_2}\lambda^+_{X_3X_2}) - k_{X_2}X_2 \\
\dot{X}_3 &= G_{X_3} H^S(X_3, X_3X_3, n_{X_3X_3}, \lambda^+_{X_3X_3}) H^S(X_5, X_5X_{30}, n_{X_5X_3}, \lambda^-_{X_5X_3}) H^S(X_7, X_7X_{30}, n_{X_7X_3}, \lambda^-_{X_7X_3}) / \lambda^+_{X_3X_3} - k_{X_3}X_3 \\
\dot{X}_4 &= G_{X_4} H^S(X_5, X_5X_{40}, n_{X_5X_4}, \lambda^+_{X_5X_4}) H^S(X_7, X_7X_{40}, n_{X_7X_4}, \lambda^+_{X_7X_4}) H^S(X_9, X_9X_{40}, n_{X_9X_4}, \lambda^+_{X_9X_4}) / (\lambda^+_{X_5X_4}\lambda^+_{X_7X_4}\lambda^+_{X_9X_4}) - k_{X_4}X_4 \\
\dot{X}_5 &= G_{X_5} H^S(X_1, X_1X_{50}, n_{X_1X_5}, \lambda^-_{X_1X_5}) H^S(X_4, X_4X_{50}, n_{X_4X_5}, \lambda^+_{X_4X_5}) H^S(X_5, X_5X_{50}, n_{X_5X_5}, \lambda^-_{X_5X_5}) H^S(X_6, X_6X_{50}, n_{X_6X_5}, \lambda^+_{X_6X_5}) \\
&\quad H^S(X_8, X_8X_{50}, n_{X_8X_5}, \lambda^+_{X_8X_5}) H^S(X_3, X_3X_{50}, n_{X_3X_5}, \lambda^-_{X_3X_5}) / (\lambda^+_{X_4X_5}\lambda^+_{X_6X_5}\lambda^+_{X_8X_5}) - k_{X_5}X_5 \\
\dot{X}_6 &= G_{X_6} H^S(X_5, X_5X_{60}, n_{X_5X_6}, \lambda^+_{X_5X_6}) / \lambda^+_{X_5X_6} - k_{X_6}X_6 \\
\dot{X}_7 &= G_{X_7} H^S(X_2, X_2X_{70}, n_{X_2X_7}, \lambda^-_{X_2X_7}) H^S(X_3, X_3X_{70}, n_{X_3X_7}, \lambda^-_{X_3X_7}) H^S(X_8, X_8X_{70}, n_{X_8X_7}, \lambda^+_{X_8X_7}) / \lambda^+_{X_8X_7} - k_{X_7}X_7 \\
\dot{X}_8 &= G_{X_8} H^S(X_7, X_7X_{80}, n_{X_7X_8}, \lambda^+_{X_7X_8}) H^S(X_9, X_9X_{80}, n_{X_9X_8}, \lambda^+_{X_9X_8}) / (\lambda^+_{X_7X_8}\lambda^+_{X_9X_8}) - k_{X_8}X_8 \\
\dot{X}_9 &= G_{X_9} H^S(X_8, X_8X_{90}, n_{X_8X_9}, \lambda^+_{X_8X_9}) / \lambda^+_{X_8X_9} - k_{X_9}X_9
\end{aligned}
\quad , (27)
$$

where G_i represents the maximum production rate of gene i, and k_i represents the degradation rate of gene i. For the regulatory link (either activation or inhibition) from gene i to gene j, ij_0 represents the threshold level of gene i, n_{ij} is the Hill coefficient, and λ_{ij} is the maximum fold change of the gene j level caused by gene i. $\lambda_{ij} < 1$ for an inhibitory regulation, referred to as λ^-_{ij}, while $\lambda_{ij} > 1$ for an excitatory regulation, referred to as λ^+_{ij}. X_n (n = 1, 2 … 9) stand for the expression levels of *Gata6*, *Gcnf*, *Cdx2*, *Klf4*, *Nanog*, *Pbx1*, *Oct4,* OCT4-SOX2 complex and *Sox2*, respectively. In this model, the OCT4-SOX2 complex is modeled as a transcription factor.

Through RACIPE method, we generated 10,000 sets of parameters (10,000 RACIPE models) for this circuit. To find the stable steady states for each RACIPE model, we simulated the above ordinary differential equations (ODEs) using the Euler method until the system reaches to a steady state. To explore the solution space thoroughly, we repeated the above calculations 1,000 times with different initial conditions. From the 1,000 steady state solutions, we identified the number of distinct

stable states and the expression of genes for each stable state. The protocol was tested on the stem cell circuit with 10 independent runs, and we identified the converged probability distribution of the number of stable steady states for each model and the probability distribution of the identified gene states (Figure S3.1). We found some rare occasions (<1%) to have a RACIPE model with more than six coexisting stable steady states. Since they are not statistically significant, we excluded these data from further analysis. The circuit also has ~ 2% chance to have oscillatory or chaotic dynamics, and we also excluded them from the analysis.

Normalization of gene expression data

From an ensemble of the RACIPE models, we collected a large set of gene expression data, very similar to those obtained in experiment. Therefore, we can apply tools for data analysis to these computationally generated data. Here, we processed the computational expression data by a standard normalization method before we performed clustering analysis. The gene expression levels from the WT models are normalized by first log transformation and standardization, i.e.

$$x_i^{WT} \to \frac{\log_2(x_i^{WT}) - \overline{\log_2(x_i^{WT})}}{\sigma(\log_2(x_i^{WT}))}, \quad (28)$$

The gene expression levels from the treatment models is normalized by first log transformation and standardization by the mean ($\overline{\log_2(x_i^{WT})}$) and the standard deviation ($\sigma(\log_2(x_i^{WT}))$) of the WT models

$$x_i^{T} \to \frac{\log_2(x_i^{T}) - \overline{\log_2(x_i^{WT})}}{\sigma(\log_2(x_i^{WT}))}, \quad (29)$$

For the single cell gene expression data of mouse embryo cells[117], we followed the procedures in the original paper for data normalization.

Clustering analysis

We performed clustering analysis on the normalized gene expression data from the RACIPE models as shown in Figure 4.1b (left). Each column represents a gene, and each row represents a stable steady state of the circuit for a particular RACIPE model. For RACIPE models with more than one stable state, the gene expression profiles for all of the stable states were shown in multiple rows. We applied the average linkage hierarchical clustering analysis using Euclidean distance by Cluster 3.0[152] and the results were visualized by JavaTreeview[153]. The cutoff distance was chosen to be 0.8 so that major clusters have probability more than 0.5% and each cluster has a distinct gene expression pattern. These clusters correspond to different gene states for the circuit. Principal component analysis was performed on the same data by using "pca" function in Matlab/2014b (Figure 4.1d). The major gene clusters can be readily recognizable from the probability density map projected onto the first two principal component axes.

The single cell gene expression data of mouse embryo cells[117] were analyzed by the same average linkage hierarchical clustering analysis with Euclidean distance, from which we identified the same nine clusters (Figure S3.5) as those found in the original paper. To compare the experimental data with the predicted data from the RACIPE, we extracted the data of the 6 common genes (*Gata6*, *Cdx2*, *Klf4*, *Nanog*, *Oct4*, and *Sox2*), as shown in the second heat map (Figure 4.1b, right).

Random gene expression data

Throughout the study, we utilized randomly generated gene expression data for statistical significant test. We first computed the histogram of gene expression for each gene from all the RACIPE models. A random gene expression vector was constructed so that each element has value randomly sampled from the histogram of the corresponding gene.

Comparison between the gene expression data from the RACIPE models and the experiments. The similarity between two gene expression vectors can be quantified by the mean square error (MSE^{XY})

$$MSE^{XY} = \frac{1}{n}\left\|X - Y\right\|_2^2, \quad (30)$$

where X and Y are the gene expression vectors with the same number of genes n. The choice of X and Y will be explained in the next two paragraphs. X and Y were regarded to be significantly similar if MSE^{XY} is better than 99% MSE^{XZ}, where

$$MSE_i^{XZ} = \frac{1}{n}\left\|X - Z_i\right\|_2^2, \quad (31)$$

Z consists of 10,000 random gene expression vectors (each vector is labeled by i, see the above section).

Cluster-based analysis: An experimental gene state cluster was matched to a gene state cluster from RACIPE as follows. We calculated the MSE^{XY} between the average gene expression vector of the experimental cluster (X) and the average gene expression vector for every RACIPE cluster (Y) (Figure S3.6). Then, we calculated the MSE^{XZ}

between the average gene expression vector of the experimental cluster (X) and each of the 10,000 random gene expression vectors (Z). The experimental cluster was considered to be significantly similar to a RACIPE cluster when the MSE^{XY} value is smaller than 99% of the MSE^{XZ} values.

Individual-based analysis: Alternatively, we calculated the fraction of the RACIPE models whose gene expression vector matches any one of the experimental clusters by a one-on-one comparison. For each experimental gene state cluster, we calculated the MSE^{XY} between the average gene expression vector of the experimental cluster (X) and the gene expression vector of each stable state for every RACIPE model (Y). The RACIPE model was considered to be significantly similar to one of the experimental clusters according to the above criterion.

Null model for hypothesis test: p values were computed for the comparison between the experimental data and the RACIPE data. For the cluster-based analysis, we generated 15 random gene expression vectors to represent the average gene expression vectors of 15 gene states, and randomly assigned the probability of a different RACIPE cluster to each vector. This data serves as a null model, and they were compared with the experimental data. This procedure is repeated for 10,000 times for statistical analysis ((Figure 4.1e and Figure S3.7a). For the individual-based analysis, we performed two random tests. First, we generated the same number of random gene expression vectors as the RACIPE models (Figure S3.7b), and used them as the null model. Second, the null model is the same RACIPE models, but for each cluster the gene identities were shuffled (Figure S3.7c).

Comparison of the parameters between two gene states.

Each parameter from all of the RACIPE models was first standardized by subtracting its mean and dividing the difference by its standard deviation. Then, we found the RACIPE models (group I) that have stable states from the gene state cluster i but no states from the cluster j. Similarly, we searched for the RACIPE models (group J) that have stable states from the gene state cluster j but no states from the cluster i. The mean of each parameter in both groups was plotted in a 2D diagram (Figure 4.1d and Figure S3.8). Note that the fold changes of the inhibitory regulations were inversed before the normalization.

Comparison with random circuits

We applied the RACIPE to each random circuit. Using the same cutoff distance as the stem cell circuit, we identified major clusters (gene states) by using hierarchical clustering analysis (Figure S3.11). The clusters from the random circuits were compared with experimental data by both the cluster-based and the individual-based methods (Figure 4.2b and Figure S3.13).

We used k nearest neighbors approach[188] to estimate the local density around each data point in the high dimensional gene expression space. The local density of a point x is

$$P(x) \cong \frac{k}{NV} = \frac{k}{N \cdot c_D \cdot R_k^D(x)}, \quad (32)$$

where k is number of nearest neighbors for the density estimation (set to be 100 here), N is the total number of data points and V is the volume that encloses the k neighbors surrounding the point x. c_D is the volume of the unit sphere in D dimensions, which is equal to $c_D = \pi^{D/2} / \Gamma(D/2+1)$. $R_k^D(x)$ is the distance between the point x and its k-th nearest neighbor.

Classification of the gene expression data

By the hierarchical clustering analysis of the RACIPE models, we identified 15 major gene state clusters (Figure 4.1d, left). The gene expression data of these gene states serve as the references for later classification of an unclassified gene expression data. To do this, we assigned the unclassified gene expression profile (i) to the gene state cluster with the smallest "minimum Euclidean distance" (D_{in}) between the i and the reference gene state n. D_{in} is defined as

$$D_{in} = \min_{1 \le j \le N_n} \left(\sqrt{\sum_{g=1}^{9} (X_{ig} - R_{njg})^2} \right), \quad (33)$$

where N_n is the number of gene expression data in the gene state cluster n (n = 1, 2 … 15). X_{ig} is the expression level of gene g in the steady state i, and R_{njg} is the expression level of gene g for every steady state j in the reference gene state n.

To calculate the probability of observing each gene state cluster from all the RACIPE models (Figure 4.1c and Figure S3.1a), we computed the weighted count of all the gene expression data that belong to the cluster. We weighted each gene expression

data by the number of stable steady states for the corresponding RACIPE model. For example, if the RACIPE model is bi-stable, the contributions of both of the stable steady state data are ½ instead of 1.

Perturbation treatment

For the perturbation treatment, we generated a new ensemble of models by specifying certain parameters and randomizing the rest parameters in the same ranges as before. We simulated the knockout (KO) of a gene by setting the maximum production rate (G_i) of the gene to be zero. For the linkage removal, we set the fold change of the regulation to be 1 (no fold change). For the activation (inhibition) of a gene, we scaled the maximum production rates (G_i) by a certain level. We referred to the RACIPE models for the original circuit as WT models, and those for the circuit under certain treatment as treatment (T) models.

Hybrid RACIPE data

We randomly picked 1,000 steady states from the original RACIPE models (WT), and the ones with the treatment on each gene (except for the OCT4-SOX2 complex). The treatment includes inhibition of a gene by 50 folds and activation of a gene by 50 folds. Together, we compiled a hybrid gene expression data set with 1,000 original data and 16 times 1,000 new treatment data. By using the same hierarchical clustering analysis on this hybrid data set, we again identified very similar gene state clusters as the original RACIPE models. The comparison between the gene state clusters of the hybrid data and those of the original data was performed by the cluster-based analysis (Figure S3.9).

Analysis of knockout

To measure the effects of the gene knockout, we calculated Kullback–Leibler (KL) divergence (D_{kl}) for the knockout of each gene, defined as

$$D_{KLi}(P_{WT} \,||\, P_{iKO}) = \sum_{n} P_{WT}(n) \ln \frac{P_{WT}(n)}{P_{iKO}(n)}, \quad (34)$$

where P_{WT} is the probability distribution of the number of stable steady states for each RACIPE model for the circuit without treatment (WT), and P_{iKD} is that for the circuit with the knockout (KO) of gene i. n stands for the RACIPE models with n stable states. The KL divergence was also applied to evaluate the effects of the removal of each regulatory link (Figure 4.3b).

Population heterogeneity

We quantified the population heterogeneity by calculating the weighted information entropy (phylogenetic)[189]

$$H = - \sum_{branches\ b\ of\ T} l(b) P(b) \log_2 P(b), \quad (35)$$

where T is a tree derived from the hierarchical clustering analysis on the RACIPE models (Figure 4.1b, left), $l(b)$ is the length of a branch b of the T, and $P(b)$ represents the probability of all the gene states that are represented by the leaves descending from b. More heterogeneous population tends to have higher entropy.

4.3. Results

Stem cell circuit performs robust functions against environmental variability

For majority of the models (~98%), the circuit allows one or multiple stable steady states (Figure 4.2a and Figure S3.3a). We collected into one dataset the gene expression profiles for the stable states of all the models (Figure 4.1b). This dataset resembles experimental gene expression data, thus inspiring us to apply similar biostatistics tools to analyze the data. Using hierarchical clustering analysis (HCA, Methods), we found that these gene expression profiles form fifteen gene state clusters (Figure 4.1b, c). The clustering results were confirmed by principal component analysis (PCA) (Figure 4.1d). It is worth noting that these gene state clusters were observed even though there are large variations in the circuit kinetic parameters. Similarly, robust clusters were also observed for several simple toggle-switch-like circuit motifs and a biological gene circuit for cancer metastasis in our previous study (Figure 3.3 and Figure 3.6). These observations suggest that the topology of gene circuits define all possible gene expression patterns, and the RACIPE method is ideal to identify these gene expression patterns in an unbiased manner.

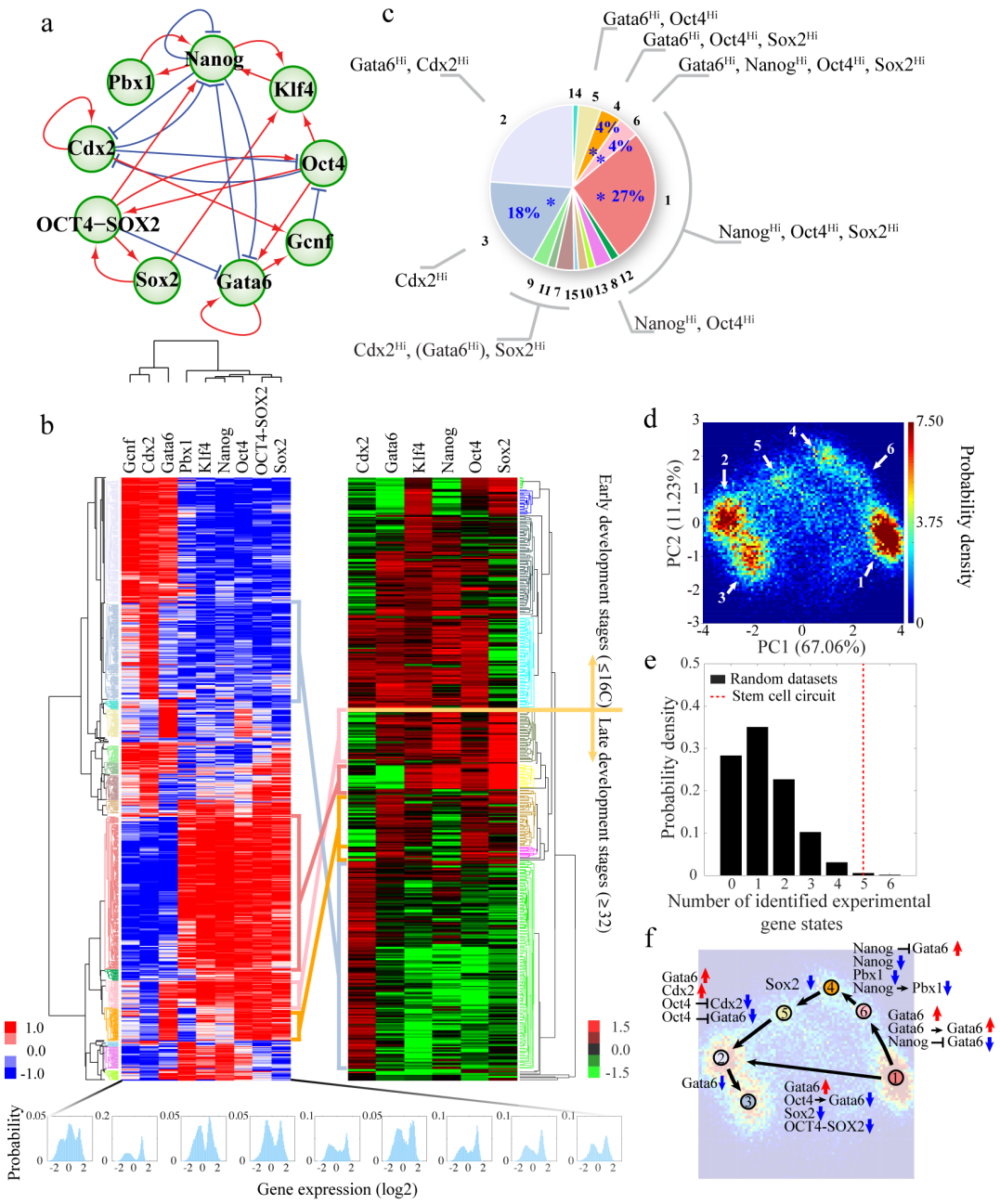


Figure 4.1 The RACIPE method uncovers robust gene states for the stem cell circuit.

(a) Diagram of the core gene regulatory circuit for stem cell differentiation. Red lines and arrows represent activations; blue lines and bars represent inhibitions. **(b)** Comparison of the gene expression of the circuit between the RACIPE prediction (left) and mouse embryo single cell data (right). In both heat maps, each column represents a

gene; each row represents the gene expressions for a stable steady state of a RACIPE model (left) or that for a single cell (right). Robust clusters (gene states, colored hierarchical trees) were identified for both data sets by unsupervised hierarchical clustering analysis. Four computationally predicted gene states match those from the late stage single cell data. The histogram of the predicted expression levels for each gene is shown at the bottom (blue, 50 bins in each histogram). **(c)** Characteristic gene expressions for each gene state, ranked by the likelihood in the RACIPE models. The four gene states that match the experimental data are highlighted by blue asterisks and are shown with their likelihoods. **(d)** 2D probability density map of the RACIPE predicted gene expression data projected to the 1st and 2nd principal component axes. The most probable gene states (top 6) are labeled. **(e)** Histogram of the number of experimental gene states identified by a random dataset (see Methods). The RACIPE (red dotted line) outperforms more than 99% of random models. **(f)** Key parameters that are involved in the transitions among certain gene states.

RACIPE predications are experimentally consistent

Next, we evaluated the power of RACIPE in predicting experimentally observed gene expression profiles. We found that the gene expression patterns for the most probable predicted RACIPE clusters have been observed experimentally in the literature (Figure 4.1c and Figure S3.1b). During embryonic development, the first cell fate determination happens at the stage of blastocyst where the inner cell mass (ICM) and trophectoderm (TE) are formed[190]. *Oct4* and *Sox2* were reported to express throughout ICM[191]. At the early differentiation of inner cell mass (ICM), *Gata6* and *Nanog* may co-express (State 6)[192,193], but *Nanog* is required together with *Oct4* and *Sox2* for cells to commit to epiblast and reach ground sate pluripotency (State 1 and 12)[191,194]. Further differentiation of mouse embryonic stem (mES) cells into mesendoderm requires *Nanog* and *Oct4* instead of *Sox2* (State 8)[195]. During the *Gata6* induction of mES cells into extraembryonic endoderm (ExE), a step-wise pluripotency factor disengagement was

reported, with initial repression of *Nanog* and *Esrrb*, then *Sox2*, and finally *Oct4*, where the co-expression of *Gata6* and *Oct4* (State 5) was observed[194]. On the other hand, down regulation of *Oct4* induces the differentiation of mES cells into trophoblast with the increase expression of *Cdx2* and *Gata6* (State 2)[196,197]. Overexpression *Cdx2* is sufficient to generate proper trophoblast stem (TS) cells (State 3)[117,198]. TE could further differentiate into ExE, where *Cdx2*, *Gata6* and *Sox2* may be all expressed (State 7,9, 11 and 15)[198–200].

In addition, we compared the RACIPE results with single cell gene expression data for mouse embryo cells at different development stages[173] (Figure 4.1b). Interestingly, the RACIPE gene states only match those from the late stage of embryo development (≥32 cells), where totipotent cells start to differentiate into trophectoderm (TE) and inner cell mass (ICM), but not from the early stage (≤16 cells) (Figure 4.1b), suggesting that this specific core circuit may be only operative at the late stage. These matched gene states - the $Cdx2^{Hi}$ state, the $Gata6^{Hi}/Nanog^{Hi}/Oct4^{Hi}/Sox2^{Hi}$ state, the $Nanog^{Hi}/Oct4^{Hi}/Sox2^{Hi}$ state and the $Gata6^{Hi}/Oct4^{Hi}/Sox2^{Hi}$ state – are among the most probable gene states from the RACIPE models (Figure 4.1b, c). Notably, the good matching to the experimental data is statistically significant (Figure 4.1e and Figure S3.7).

Therefore, similar as single cell gene expression data measured by qPCR or RNA-seq, RACIPE could predict the expression of genes in the circuit based on its topology information, subject to the same analysis for the experimental data to infer the information of subpopulations and characteristic expressed genes in each subpopulation.

However, more than statistical inference from the data, RACIPE meanwhile established the physical model of the circuit underlying the data, thus enable us to uncover the fundamental design principles of the circuit functions. For example, from the RACIPE models, we can identify the parameters that are significantly changed between two gene states (Figure 4.1f and Figure S3.8). These parameters could be the best candidates to target to change during cell phenotypic transitions.

Stem cell circuit is well evolved circuit to perform robust functions

To further investigate the role of the circuit topology to its functions, we constructed two types of random circuits (Figure S3.10 and Methods) with similar structures as the stem cell circuit. Both types of circuits (Figure S3.10) preserve the total number of regulatory links and genes, the number of inward links and the number of outward links for each gene, and the activations from *Oct4*, *Sox2* to OCT4-SOX2. For the Type I random circuit (Figure S3.10a), in addition, we kept the same number of excitatory inward links and inhibitory inward links. Compared to the stem cell circuit, the random circuits are much more likely to generate oscillatory or chaotic dynamics, and are more likely to have only one stable state for each RACIPE model (Figure 4.2a). When the gene states from all the RACIPE models are combined, the expression histogram for each gene typically has multiple peaks for the native stem cell circuit (blue in Figure 4.2c), but has only a single peak for many genes for the random circuits (red in Figure 4.2c). We also found that it is difficult to cluster the gene expression data in the case of the random circuits (Figure S3.11), partly because the stem cell circuit has much higher local density of the RACIPE gene expression data than a random circuit (Figure 4.2d and

Figure S3.12). Furthermore, the RACIPE models of the stem cell circuit match the mouse embryo single-cell data significantly better than those of a random circuit (Figure 4.2b and Figure S3.13). These findings suggest that the stem cell circuit has evolved to be robust to perform its biological function and that RACIPE can reveal that function.

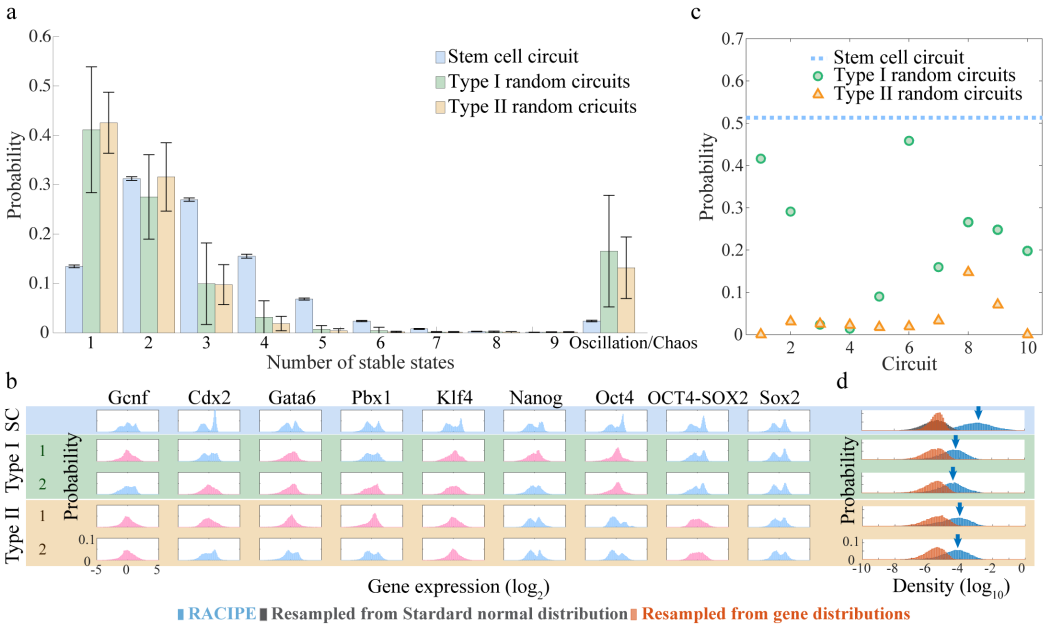


Figure 4.2 Comparison between the stem cell circuit and a random circuit.

(a) Probability distribution of the number of stable steady states of each RACIPE model for the stem cell circuit and random circuits (type I: Figure S3.10a, type II: Figure S3.10b, and see Methods). The error bar for the stem cell circuit shows the standard deviation for 10 simulation repeats, while the ones for the random circuits represent the standard deviations for the 10 random circuits of each type. **(b)** Percentage of the RACIPE predicted gene expression data in consistent with the experimental single cell gene expression data in Figure 4.1b. **(c)** Histogram of the expression for each gene for the stem cell circuit and the random circuits. The multi-peak distributions are in blue and the single peak distributions are in red. The columns represent genes and rows correspond to different circuits. Two representative random circuits for each type are presented, and the rest results are shown in Figure S3.12. **(d)** Histogram of the local density distributions of the RACIPE predicted gene expression data (blue, mean value is highlighted by blue downward arrows), the random gene expression data resampled from the RACIPE

expression distribution of each gene (orange), the random gene expression data resampled from standard normal distributions (grey).

Hierarchical structure in stem cell circuit

In addition to identifying the gene states, RACIPE can also be applied to decode the design principles of a gene circuit, as shown in the following example. We can use perturbation analysis (gene knockout (KO) and link removal from the RACIPE analysis) to identify the most important genes and regulatory links to the circuit dynamics. In the case of the stem cell system, we found *Oct4*, *Sox2*, OCT4-SOX2 and *Cdx2* to be most sensitive to the knockout treatments (Figure 4.3a-c), as measured by the changes of the probability distribution of the number of stable states for each RACIPE model (defined by the KL divergence, see Methods). These gene components form the first decision-making sub-circuit (Figure 4.3d, top layer, *Oct4*/*Cdx2*), and the rest of the components form the second one (bottom layer, *Gata6*/*Nanog*). Interestingly, the dynamic behavior of the upper sub-circuit is much less disrupted than that of the lower sub-circuit when the regulatory links connecting the two sub-circuits are removed (Figure S3.14). Our findings are consistent with our current knowledge of mouse embryo development[201], where Oct4/Cdx2 work as the first decision gate for inner cell mass and trophectoderm differentiation while *Gata6/Nanog* specify the lineage between epiblast and primitive endoderm.

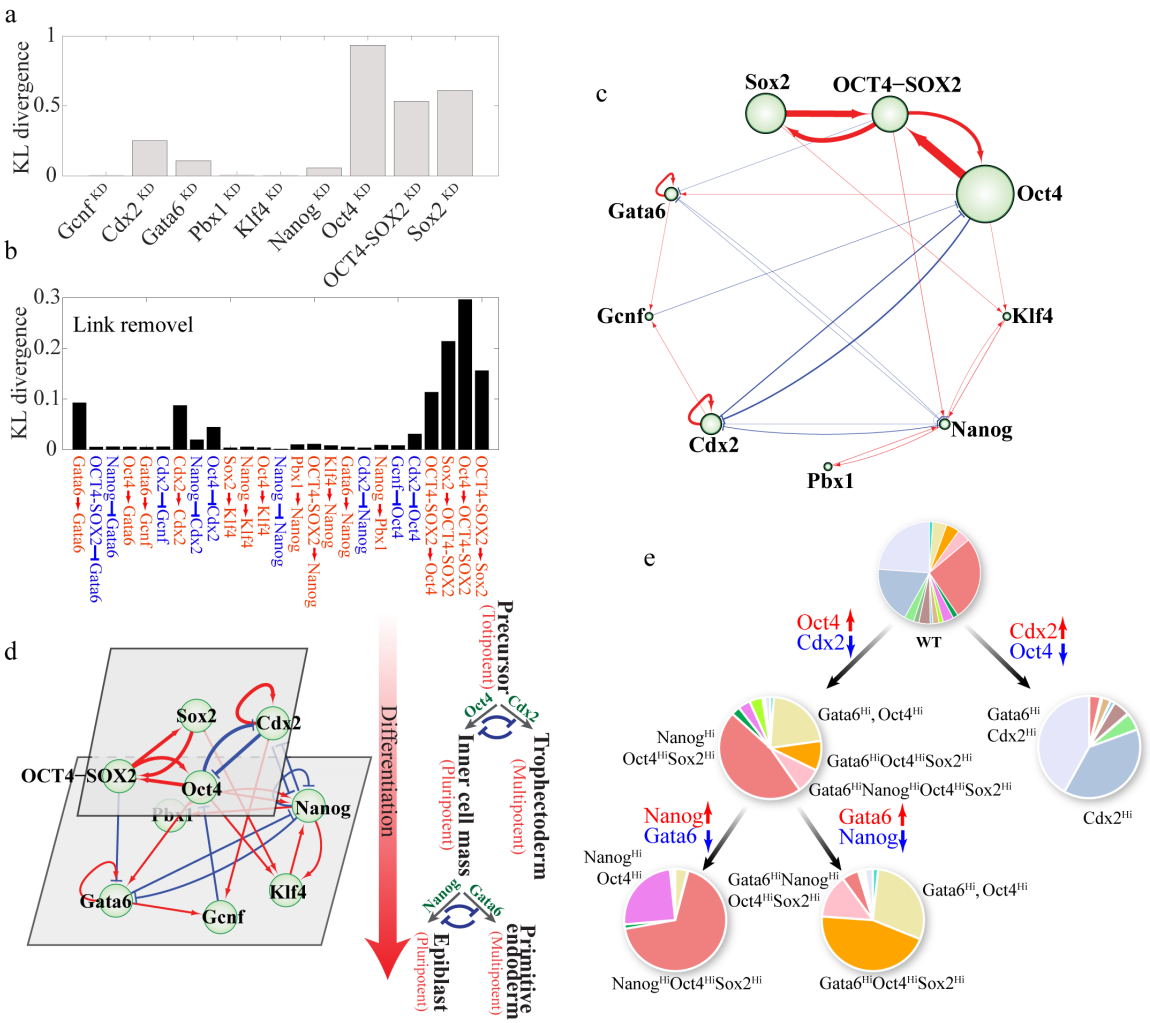


Figure 4.3 Hierarchical structure of the stem cell circuit inferred from the perturbation analysis.

(a) The Kullback-Leibler (KL) divergence between the probability distributions of the number of stable states for each RACIPE model computed before and after the knockout (superscript KO) of each gene. **(b)** Similar to (a), but the KL divergences are between the distributions before and after removal of each regulatory link. **(c)** Schematic diagram of the stem cell circuit highlighting the important genes and regulatory links. The larger the gene element and the ticker the regulatory link, the more importance the component to the circuit behavior, as inferred from the analyses in (a) and (b). **(d)** The hierarchical structure of the stem cell circuit (left) is consistent with the two-step decision-making of mouse embryonic development (right). **(e)** The roadmap of stem cell differentiation inferred from the RACIPE simulations. All the original RACIPE models (WT) were treated by activating (↑) or inhibiting (↓) the production of the corresponding genes. The

probability distribution of the gene states for each condition was plotted in the corresponding pie chart.

Understand the population behavior of stem cells

We further explored the possibility of using the ensemble of the RACIPE models to simulate the behavior of a cell population. Indeed, the idea of kinetic parameter variation has been used to model the expression of an inducible gene in an yeast population[143]. RACIPE allows us to extend this idea to the stem cell population. Thus, the roadmap for the stem cell development can be explained by starting from a population of precursor cells modeled by the original RACIPE models and stimulating the corresponding gene expressions in each RACIPE model by a relevant external signal (Figure 4.3e and Methods). We found that strong activation of *Cdx2* converts most RACIPE models to the $Cdx2^{Hi}$ states (trophectoderm)[202], while strong activation of *Oct4* converts most models to either the $Nanog^{Hi}$ or the $Gata6^{Hi}$ states (inner cell mass)[201,203]. After strong activation of *Oct4*, additional signaling effects on *Nanog* and *Gata6* converts cells to either the $Nanog^{Hi}$ state (epiblast) or the $Gata6^{Hi}$ gene state (primitive endoderm)[201] (Figure 4.3e).

Within this scenario, we propose to quantify the population heterogeneity in the cell culture of stem cells[172,204] by the information entropy of the RACIPE models[189]. In the simulation, we then added an activating or inhibiting signal on each gene, and evaluated the changes in the entropy (Figure S3.15). Either activation or inhibition of a gene usually decreases the entropy, resulting in a more homogeneous population, typically observed during cell differentiation. This may explain why some pluripotent genes, such as *Oct4*, can commit differentiation with either high- or low-

expression[205]. Also, we found that the decrease in the entropy is most pronounced with the inhibition of *Oct4*/*Sox2* and the activation of *Cdx2*, consistent with view of these genes as determining stem cell differentiation. All of these findings suggest the possibility of using the ensemble of the RACIPE models to quantify and understand population behaviors.

4.4. Discussion

From RACIPE analysis, we demonstrated that the core regulatory circuit of stem cell differentiation, as a well evolved circuit in topology, encodes the robust functions against the cell-cell variability caused by internal (gene mutations, low copy number of transcripts and transcriptional bursting etc.) and external (epigenetic states and signaling states etc.) noises. Moreover, we revealed that stem cell circuit has a hierarchical structure, in consistent with the two-step differentiation process for mouse embryo development. We expect RACIPE to be an unbiased computational approach to predict and decipher the functions of many other gene circuits or even large gene networks.

The integration of all the findings from the above analyses seem to provide a new picture of stem cell differentiation beyond the traditional Waddington's epigenetic landscape to explain how genes regulate development[206]. Waddington's epigenetic landscape is a popular metaphor, which considers the differentiation of progenitor cells is analog to the process of marbles rolling along a landscape, in which cell potency decreases from hill to valleys during stem cell differentiation. However, the exact meaning of this epigenetic landscape remains elusive. Until now, the most common

views[184,186] map each of the differentiated phenotypes or each of the progenitor cells of different cell potency to a certain cell state with unique stable gene expression profiles, and the epigenetic landscape can be quantified by an effective potential[184,186]. Thus, cell differentiation was regarded as the transition from the progenitor cell state to the differentiated cell states.

Here we propose a possible different interpretation of Waddington's epigenetic landscape. A stem cell phenotype, instead of being a specific cell state, is regarded as a heterogeneous population of cells in various states, each of which corresponds to a differentiated lineage with a distinct gene expression pattern[207]. Cells with high cell potency (i.e. the capability of stem cells to differentiate into other cell types) are plastic to convert into the various cell states stochastically by both the intrinsic factors (gene expression noises, fast process) and the extrinsic factors (transient epigenetic regulation and cell signaling, slow process). However, when cells are subject to stable perturbations, they lose the capacity to access certain cell states, therefore being less heterogeneous, i.e. smaller informational entropy, and differentiated. Our view is consistent with the following evidences from both experiments and our RACIPE analysis on the stem cell circuit. First, the stem cell progenitors of either totipotency or pluripotency have highly heterogeneous gene expression, and several cell sub-populations of differentiated types, called lineage priming, have been identified in cell culture[208–210]. In our model, we also identified fifteen distinct cell phenotypes by the RACIPE analysis on the core stem cell circuit (Figure 4.1b, left), some of which became inaccessible during the *in silico* differentiation. Second, some pluripotency factors[207], including Oct4, commit differentiation when they are either down- or over-

expressed[205]. With RACIPE, we showed that the heterogeneity of cell population drops dramatically when down-regulating Oct4, while it is at the maximum when Oct4 is expressed at intermediate level (Figure S3.15). Therefore, we might associate the population heterogeneity with stem cell potency.

One limitation to the current RACIPE method is how to accurately construct the core biological circuit for the certain function, although the master regulators in the core circuit often have been widely studied in experiments. The core circuit for stem cell differentiation here is constructed based on extensive literature search, which might miss out some important regulatory links or genes. One practical way to verify the reliability of the core circuit is to compare the predicted gene expression data with the experimental ones, which may also in turn suggest the missing of the links or genes. Also, the comprehensive combination of genomics data such as ChIP-Seq maps and biochemistry experiments[211] may be able to generate an experimentally validated gene circuit subject to RACIPE analysis. However, it still remains a big challenge to build a large gene regulatory network in accurate.

To conclude, we showed from the results of parametric randomization on a proposed stem cell circuit, that the circuit topology plays the essential role in determining the gene expression patterns of cells. This result is analogous to findings from protein structure modeling, where conformational motions have been found to be determined by the overall molecular shape[212] and protein folding process by native residue contacts[213]. RACIPE allows the interrogation of the circuit behavior by parametric randomization, from which we can learn about its robustness, heterogeneity and function.

RACPE is expected to facilitate the discovery of the causal relationship between the gene regulatory networks and experimental gene expression data.

Chapter 5

Modeling putative therapeutic implications of exosome exchange between tumor and immune cells

Development of effective strategies to mobilize the immune system as a therapeutic modality in cancer necessitates a better understanding of the contribution of the tumor microenvironment to the complex interplay between cancer cells and the immune response. Recently, effort has been directed at unraveling the functional role of exosomes and their cargo of messengers in this interplay. Exosomes are small vesicles (30–200 nm) that mediate local and long-range communication through the horizontal transfer of information, such as combinations of proteins, mRNA and microRNAs. Here, we develop a tractable theoretical framework to study the putative role of cell to cell signaling through exosome exchange on the cancer-immunity interplay. We reduce the complex interplay into a generic model whose three components are cancer cells, dendritic cells (consisting of precursor, immature and mature types) and killer cells

(consisting of cytotoxic T cells, helper T cells, effector B cells and natural killer cells). The framework also incorporates the effects of exosome exchange on enhancement/reduction of cell maturation, proliferation, apoptosis, immune recognition and activation/inhibition. We reveal co-occurrence of three cancer states: low risk cancer with intermediate immunity (L), intermediate risk cancer with high immunity (I) and high risk cancer with low immunity state (H), and establish the corresponding effective landscape for the cancer-immunity network. We illustrate how the framework can contribute to the design and assessment of combination therapies.

5.1. Introduction

Immunotherapeutic approaches have recently emerged as effective therapeutic modalities[214] exemplified by immune checkpoint blockade with anti-CTLA-4 to activate T-cells and induce tumor cell killing which has been shown to be effective for some cancers but not others[215]. A better understanding of the intricate interplay between cancer and the immune system, and of mechanisms of immune evasion and of hijacking of the host response by cancer cells is relevant to the development of effective immunotherapeutic approaches [216–219].

The immune based suppression of tumor development and progression is mediated through non-specific innate immunity and antigen-specific adaptive immunity [220]. On the other hand, cancer cells can inhibit the immune response, thus evading suppression in multiple ways [221] (see below for details), and additionally hijack the immune system to their advantage [216,217]. The challenge to understand the tumor-

immune interplay stems from the dynamic nature, and complexity and heterogeneity of both the cancer cells and the immune system and their interactions through the tumor microenvironment [222].

Here we consider immune cells as consisting of macrophages [223], natural killer cells [224], cytotoxic T cells [225], helper T cells [226], and regulatory T cells [216]. These various immune cells are produced, activated and perform their functions separated by space and time, which contributes additional complexity [227]. Among the immune cells, dendritic cells (DCs) are the most efficient antigen presenting cells (APCs) to bridge innate immunity with adaptive immunity [228]. DCs also secrete cytokines that promote the antitumor functions of both nature killer cells and macrophages [229,230]. We consider the tumor microenvironment (TME) as comprised of a heterogeneous population of cancer cells [231], stromal cells [232] and tumor infiltrating immune cells [233]. The interactions among these cell types contribute to tumor development and progression. Tumor-associated macrophages (TAM) and cancer-associated fibroblasts (CAF) regulate tumor metabolism and engender an immune-suppressive environment by secreting TGF-β and other cytokines [234]. Fluctuations in energy sources and oxygen within a tumor contribute to malignant progression and cell phenotypic diversity [235,236].

While secreted factors play critical roles in cell-cell communications, here we focus on the additional role of cell-cell communication mediated by the exchange of special extracellular lipid vesicles called exosomes [237]. These nanovesicles of about 30-200 nm are formed in the multivesicular bodies (MVBs) and then released from the

cell into the extracellular space [238]. The exosomes carry a broad range of cargo, including proteins, microRNAs, and mRNAs and DNA fragments, to specific target cells at a remote location [239]. Membrane markers assign the exosomes to specific targeted cells. Notably, upon entering the target cell, the exosomes induce modulation of cell function and even identity switch (phenotypic, epigenetic and even genetic) [240]. Exosomes have recently emerged as playing an important role in the immune system interaction with tumors [241,242]. Tumor-derived exosomes can promote metastatic niche formation by influencing bone marrow-derived cells (BMDCs) toward a pro-metastatic phenotype through upregulation of c-Met [242]. DCs have been shown to induce tumor cell killing through release of exosomes which contain potent tumor-suppressive factors such as TNF and through activation of natural killer cells, cytotoxic T cells and helper T cells [237,243–245]. On the other hand, tumor-derived exosomes (Tex) can directly inhibits the differentiation of DCs in bone marrow [246], which strongly inhibit the dendritic-cell-mediated immune response to the tumor. In addition, Tex can also directly inhibit natural killer cells [247].

Mathematical models have been devised to study the complex interactions of cancer and immune system, including those that consider spatial heterogeneity (as reviewed in ref. [248]) and those that consider only spatially homogeneous populations (as reviewed in ref. [249]). Cancer-immunity models have been constructed to investigate the effects of therapy [250–252], cancer dormancy [253] and interactions with time delay [254]. Other types of modeling methods have also been applied. For example, tumor growth has also been fitted to experimental data by artificial neural networks

[255]; a detailed network of cancer immune system has been modeled by multiple subset models [256].

In this study, we have developed an exosome-mediated (EM) model, to incorporate the special role of DCs and exosome-mediated communications. Distinct from the previous approaches, our modeling strategy is adapted from methodology used in studies of gene regulatory circuits, allowing us to check the multistability features of the system [257]. We find that, by including exosome exchange, the cancer-immunity interplay can give rise to three quasi-stable cancer states, that may be associated with the "elimination/equilibrium/escape phases" proposed in the immunoediting theory [258]. The EM model is also capable of explaining tumorigenesis by considering the time evolution of immune responses. Guided by the treatment simulations, we assess the effectiveness of various therapeutic protocols with and without time delay and noise.

5.2. Methods

Exosome-mediated (EM) modeling approach

Here, we develop a minimal yet workable theoretical framework for modeling the cancer-immunity interplay. The concept is to devise EM models that incorporate three generic (coarse-grained) components, i.e. the effective cancer cells (C), the dendritic cells (D) and the killer cells (K). C represents the cancer cells that actively interact with the immune system, e.g. those at the surface of a solid tumor. D represents precursor, immature and mature DCs that present tumor-associated antigens. The cytotoxic T cells,

effector B cells, helper T cells and natural killer cells are lumped together and referred to as killer cells (K), since as a group, they inhibit cancer cells after being activated by DCs.

To elucidate the effects of cell communications among these representative cell types, the dynamics of C, D and K cells are modeled by nonlinear ordinary differential equations. We consider two communication networks as illustrated in Figure 5.1. For both cases, the C population self-activates by means of exosome-mediated cell proliferation [259]. The D population self-activates by means of the cell proliferation of dendritic cells. D activates proliferation of K [243–245], while K induces maturation of D [260]. Also, K targets C and induces C's apoptosis [224–226,261,262] while C inhibits K by slowing down its proliferation and increasing its apoptosis [263,264]. The two networks, however, differ from each other in the interactions between C and D. In case I (Figure 5.1a), D directly kills C by exosomes [237] and/or some cytokines [265]; C induces the D maturation by presenting tumor-associated antigens on the surface when the concentration of C is low [228,266], but inhibits D differentiation by exosomes and/or cytokines when the concentration of C is above a threshold level [246,267]. The deterministic rate equations for the dynamics of C, D and K densities are given by

$$\begin{aligned}
\frac{d}{dt}C &= g_{C0}H^{S}_{CC}(C,\lambda^{+}_{CC}) - k_{C}CH^{S}_{DC}(D,\lambda^{+}_{DC})H^{S}_{KC}(K,\lambda^{+}_{KC}) \\
\frac{d}{dt}D &= g_{D0}H^{S}_{DD}(D,\lambda^{+}_{DD})H^{S}_{KD}(K,\lambda^{+}_{KD})H^{S}_{CD1}(C,\lambda^{-}_{CD1})H^{S}_{CD2}(C,\lambda^{+}_{CD2}) - k_{D}D, \quad (36) \\
\frac{d}{dt}K &= g_{K0}DH^{S}_{CK1}(C,\lambda^{+}_{CK1})H^{S}_{CK2}(C,\lambda^{-}_{CK2}) - k_{K}KH^{S}_{CK3}(C,\lambda^{+}_{CK3})
\end{aligned}$$

where

$$H^S_{CC}(C,\lambda^+_{CC}) = (1/[1+(C/CC_0)^{n_{CC}}]) + \lambda^+_{CC}((C/CC_0)^{n_{CC}}/[1+(C/CC_0)^{n_{CC}}])$$
$$H^S_{DC}(D,\lambda^+_{DC}) = (1/[1+(D/DC_0)^{n_{DC}}]) + \lambda^+_{DC}((D/DC_0)^{n_{DC}}/[1+(D/DC_0)^{n_{DC}}])$$
$$H^S_{KC}(K,\lambda^+_{KC}) = (1/[1+(K/KC_0)^{n_{KC}}]) + \lambda^+_{KC}((K/KC_0)^{n_{KC}}/[1+(K/KC_0)^{n_{KC}}])$$
$$H^S_{DD}(D,\lambda^+_{DD}) = (1/[1+(D/DD_0)^{n_{DD}}]) + \lambda^+_{DD}((D/DD_0)^{n_{DD}}/[1+(D/DD_0)^{n_{DD}}])$$
$$H^S_{KD}(K,\lambda^+_{KD}) = (1/[1+(K/KD_0)^{n_{KD}}]) + \lambda^+_{KD}((K/KD_0)^{n_{KD}}/[1+(K/KD_0)^{n_{KD}}])$$
$$H^S_{CD1}(C,\lambda^-_{CD1}) = (1/[1+(C/CD1_0)^{n_{CD1}}]) + \lambda^-_{CD1}((C/CD1_0)^{n_{CD1}}/[1+(C/CD1_0)^{n_{CD1}}])$$
$$H^S_{CD2}(C,\lambda^+_{CD2}) = (1/[1+(C/CD2_0)^{n_{CD2}}]) + \lambda^+_{CD2}((C/CD2_0)^{n_{CD2}}/[1+(C/CD2_0)^{n_{CD2}}])$$
$$H^S_{CK1}(C,\lambda^+_{CK1}) = (1/[1+(C/CK1_0)^{n_{CK1}}]) + \lambda^+_{CK1}((C/CK1_0)^{n_{CK1}}/[1+(C/CK1_0)^{n_{CK1}}])$$
$$H^S_{CK2}(C,\lambda^-_{CK2}) = (1/[1+(C/CK2_0)^{n_{CK2}}]) + \lambda^-_{CK2}((C/CK2_0)^{n_{CK2}}/[1+(C/CK2_0)^{n_{CK2}}])$$
$$H^S_{CK3}(C,\lambda^+_{CK3}) = (1/[1+(C/CK3_0)^{n_{CK3}}]) + \lambda^+_{CK3}((C/CK3_0)^{n_{CK3}}/[1+(C/CK3_0)^{n_{CK3}}])$$

where g_{C0}, g_{D0} and g_{K0} are the basal proliferation rates, k_C, k_D and k_K are the basal apoptosis rates. We quantify the effect of cell communication by $H^S_I(x,\lambda_I)$ for $x \in C,D,K$, the shifted Hill function (for the interaction labeled as I), defined as $H^-_I(x) + \lambda_I H^+_I(x)$, where $H^-_I(x) = 1/[1+(x/x_I)^{n_I}]$, $H^+_I(x) = 1 - H^-_I(x)$, and λ_I is the fold change of the interaction. λ^+_I represents activation ($\lambda_I > 1$), and λ^-_I represents inhibition ($\lambda_I < 1$). The model parameters are listed in the Table S4.1. In case II (Figure 5.1b), we exclude exosome-mediated interactions, so the most prominent interaction between C and D is the activation of D by recognizing antigens that are associated with C. We remain using Equation (36) to model the dynamics of the network, but change the parameters for the modified interactions, i.e. the fold changes $\lambda^+_{DC} = 1$ and $\lambda^-_{CD1} = 1$.

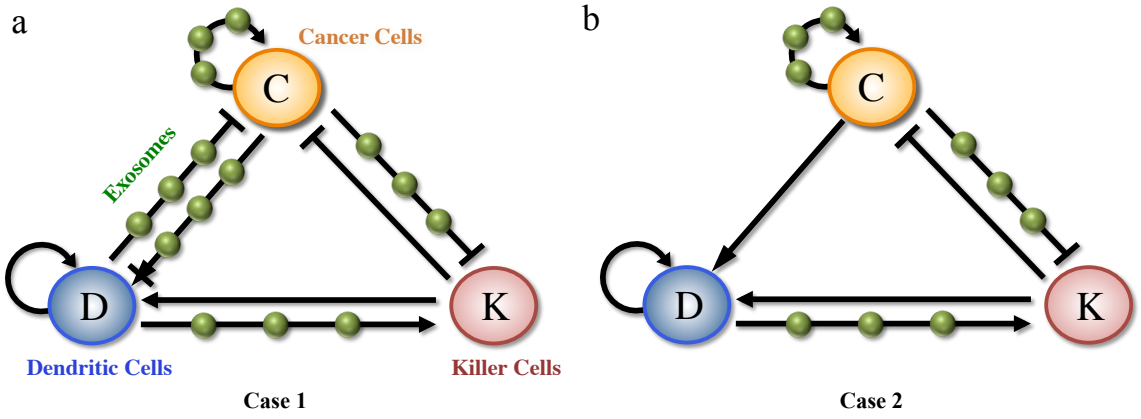


Figure 5.1 Illustration of the EM model for cancer-immune interplays.

The coarse-grained network model contains three major components: the effective cancer cells (C), the dendritic cells (D) and the killer cells (K). The links among different cell types represent the effects of cell-cell communication. An arrow denotes activation; a bar denotes inhibition; an arrow plus a bar from C to D represents activation when the C population is small, and inhibition when the C population is large. The green spheres represent cell communications that are partially mediated by exosomes. Panel a shows the full model (case I); panel b shows the model without exosome-mediated interactions between C and D (case II).

The effects of immune recognition are incorporated by varying the interactions from D to C and K to C using a recognition parameter (ρ) as follows:

$$\begin{aligned} \lambda_{DC}^{+} &= 1+\rho(\lambda_{DC}^{0}-1) \\ \lambda_{KC}^{+} &= 1+\rho(\lambda_{KC}^{0}-1) \end{aligned}, \quad (37)$$

where ρ represents the effective immune recognition: ρ = 0 corresponds to no recognition so that there is no effect from D or K to C, i.e. λ_{DC}^{+} and λ_{KC}^{+} equal to one. ρ = 1.0 corresponds to full recognition so that λ_{DC}^{+} and λ_{KC}^{+} are equal to a basal strength λ_{DC}^{0}

and λ_{KC}^{0} respectively. Note that in principle ρ can also be larger than one, which represents the case of enhanced immune response above the basal level that can be induced for example by immunotherapy.

Stability analysis

The dynamics of the EM model is described by a set of three-dimensional nonlinear rate equations Equation (36). To compute the steady state solutions, we calculate two nullclines – the first one satisfies dD/dt = 0 & dK/dt = 0 (solid navy line, Figure 5.2), and the second one satisfies dC/dt = 0 & dK/dt = 0 (solid brown line, Figure 5.2). The nullclines are projected onto the phase plane, constructed by the concentrations of the D (x-axis) and C (y-axis). The intersections of these two projected lines are the steady states for the whole system. The stability of the steady states can be further determined by the linear approximation method. The nullclines are computed by contour-based method, and the bifurcation diagrams are calculated by PITCON7 package [22].

Simulation with time delay

We consider the time delays caused by the exosome-mediated communications between the C and D cell populations. Instead of directly using Equation (36), we replace the terms $H_{DC}^{S}(D,\lambda_{DC}^{+})$, $H_{CD1}^{S}(C,\lambda_{CD1}^{-})$ and $H_{CD2}^{S}(C,\lambda_{CD2}^{+})$ by $H_{DC}^{S}(D^{*},\lambda_{DC}^{+})$, $H_{CD1}^{S}(C^{*},\lambda_{CD1}^{-})$, and $H_{CD2}^{S}(C^{*},\lambda_{CD2}^{+})$ respectively, where for $X \in C,D$ and time delay τ, X^{*} at time t is

$$X^*(t) = \begin{cases} X(0) & t \le \tau \\ X(t-\tau) & t > \tau \end{cases} . \quad [38]$$

The delay differential equations are integrated by dde23 from MATLAB [268].

Stochastic simulation

Equation (36) can be rewritten in a vector form as $\dot{\mathbf{x}} = f(\mathbf{x})$, where $\mathbf{x} = (C, D, K)^T$. In this study, we consider Gaussian white nosie. So the dynamics can be described by a Langenvin equation $\dot{\mathbf{x}} = f(\mathbf{x}) + \mathbf{L}(\mathbf{x}, t)$, where $\left\langle L_i(\mathbf{x}, t) L_j(\mathbf{x}, t') \right\rangle = 2\Upsilon\delta(t - t')$ for $i, j \in C, D, K$, Υ is a constant representing the noise level. The stochastic differential equation is integrated by Euler-Maruyama method [269]. The effective landscape for the state $\mathbf{x}$ is defined as $E = -\ln(P(\mathbf{x}))$ [96,97], where $P(\mathbf{x})$ is the probability for the system in state x.

Tumorigenesis and Treatment Modeling

We specifically model the time evolution of the level of immune recognition ρ for the tumor onset process. ρ follows

$$\rho = 1 - e^{-t/\tau} , [39]$$

where τ is the half time for developing the full immune recognition. ρ initially starts at zero, and increases until it saturated at one. The tumor onset is simulated by both Equation (36) and Equation (39), and the initial condition for (C, D, K) is (0, 100, 0) (cells/μL).

In the simulation of the radiation therapy, additional terms $-k_R C$, $-k_R D$ and $-k_R K$ are included to the three equations in Equation (36) respectively, where k_R is the cell death rate of all these cell types by the radiation therapy. In the EM model, k_R is treated to be proportional to the dose strength. In the simulation of DC immunotherapy, an additional constant term g_I is included to the D equation in Equation (36).

5.3. Results

Co-occurrence of multiple cancer states during tumorigenesis

We first consider two cases whereby case I includes exosome exchange and case II excludes exosome exchange. The quasi-steady state solutions of Equation (36) are obtained by the phase plane analysis (see Method section) for these two cases. In case I, when ρ equals 0.2, the interplay yields only one stable steady state, corresponding to high cancer/low immune (denoted as the H state, Figure 5.2a). That is because the effect of the immune system is not strong enough to limit cancer progression. When ρ equals 1.0, the interplay can give rise to three co-occurring quasi-stable steady states Figure 5.2b), corresponding to the (H) state, to intermediate cancer/high immune (denoted as the I state), and to low cancer/intermediate immune (denoted as the L state). At this stage, the strength of the effect of the cancer on the immune system and of the immune system on cancer are comparable, therefore allowing the co-occurrence of multiple states. The relative stability of each state can be inferred from the stochastic simulations (see Method section). When ρ=2.0, the immune system outweighs cancer, allowing only the low cancer/low immune state deterministically. It could, however, have some rare chances to

switch to the high cancer/low immune state stochastically (Figure 5.2). Notably, the predicted co-occurrence of three quasi-stable states in the EM model is likely related to the clinical classification of the equilibrium stage during tumor progression as defined in the immunoediting theory [258], which considers both the host-protection and tumor-sculpting action of immune system on tumorigenesis.

In case II, the interplay does not include negative interactions between C and D. These interactions are mainly mediated by the exosome exchange. As a consequence, case II dynamics are significantly different from those of case I. When ρ equals 0.2, the interplay leads to the high cancer/high immune state (Figure 5.2c). However, when ρ equals 1.0, the interplay leads to bistability – co-occurrence of the high cancer/high immune state and the low cancer/intermediate immune (L) state, meaning that the (I) state does not occur (Figure 5.2d). When ρ equals to 2.0, the interplay gives rise to the low cancer/intermediate immune state in both the deterministic and the stochastic analyses (Figure S4.1).

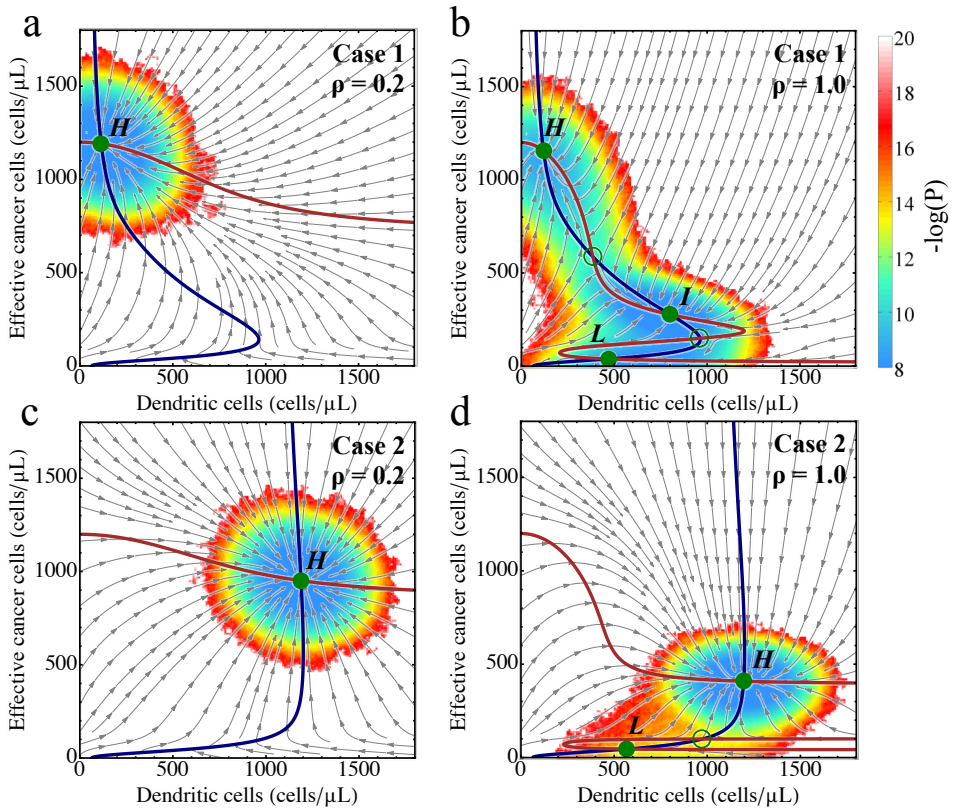


Figure 5.2 Phase plane analysis for the EM model.

In each case, a phase plane is constructed by the concentration of the dendritic cells (D, x-axis) and the effective cancer cells (C, y-axis). The nullcline dD/dt = 0 & dK/dt = 0 is shown in solid navy line and dC/dt = 0 & dK/dt = 0 is shown in solid brown line. The intersections of these two nullclines are the steady states, represented by solid green circles for stable states and by hollow green circles for unstable states. The grey arrows represent the vector field in a stream plot. The background colors illustrate the values of the effective landscape (-log(P)) computed by stochastic simulations of the network with white Gaussian noise. The states are most probable at the blue regions, and are least probable at the red/white regions. Panel a and b are for the case I network, and panel c and d are for the case II network. The immune recognition is at a low level (with value ρ = 0.2) for a and c, and is at the basal level (with value ρ = 1.0) for b and d. For the full model and ρ = 1.0 (panel b), the network allow three stable steady states – high cancer/low immune (H), intermediate cancer/high immune (I), and low cancer/intermediate immune (L).

We also show a comparison with the stability analysis for two additional cases in which a part of exosome exchange interactions are not included (Figure S4.2). The comparison between these different cases implies that exosome exchange facilitates tristability – the co-occurrence of the (H), (I) and (L) states. As we show next, the occurrence of the (I) state, which is possible due to the exosome exchange, has very important implications regarding the effectiveness of therapeutic strategies.

Bifurcation diagram for immune recognition

To further assess the effect of immune recognition in case I, we calculate and investigate the bifurcation diagrams of the cancer-immunity interplay with respect to the level of ρ as a control parameter (Figure 5.3). As is depicted in this figure, the bifurcation diagram exhibits four different regimes (phases), each corresponding to the co-occurrence of different quasi-stable cancer states: 1. a phase in which only the (L) state occurs (denoted as {L}, blue); 2. a phase in which the (L) and (H) states co-occur (denoted as {L,H}, yellow); 3. a phase in which the (L), (I) and (H) states co-occur (denoted as {L,I,H}, green); 4. a phase in which only the (H) state occurs {denoted as {H}, red}. Further inspection reveals that when ρ is above 1.1, the (L) state is the dominant one; when ρ is between 0.84 and 1.1, all three states co-occur, making the transitions among these states more likely; when ρ is below 0.84, the (H) state is the dominant one. It is worth noting that a bifurcation curve may cross itself, because the y-axis of the bifurcation diagram is projected from a 3D phase-plane (Figure 5.3c and d).

Immune recognition and tumorigenesis

Before tumor onset, immune recognition has yet been established, i.e. ρ is zero, and the C population is close to zero and the D population is low. As the tumor starts to develop, it triggers immune recognition. The fate of tumor development is determined by the relative rate of tumor growth versus the rate of increase in immune recognition. In general, for fast growing tumors, the dynamics will lead to the (H) cancer state. More specifically, depending on how fast the immune system responds (the level of immune recognition increases), which is characterized by the mean lifetime (τ) for the development of full immune recognition (ρ = 1) (see Method section), the interplay (for the model parameters of cancer growth used here) can lead to either the (H) state (when $\tau > 77$ days), the I state (when $12 < \tau < 77$ days) or the (L) state (when $\tau < 12$ days) states. Some typical examples are shown in purple lines and circles in Figure 5.3a. Further enhancement of immune recognition causes the transition of the tumor from the (H) or the (I) states to the (L) state (downward dashed arrows labeled as 1 and 2 in Figure 5.3a), suggesting possible control of the tumor by the immune system and the putative importance of boosting the immune system during early stages of tumorigenesis.

Effect of tumor evasion on immune recognition

As is now widely recognized, cancer cells can evade immune suppression during tumor development, which causes the immune recognition to gradually decrease. In general, tumor size, as quantified by the size of the C population, increases when ρ decreases. Marked growth of C can be observed in some stages either deterministically (upward dashed arrows labeled as 3 and 4 in Figure 5.3a), or stochastically when the system overcomes a dynamic barrier of the transition between two co-occurring cancer

states. Interestingly, such marked growth of C happens within phase {L, I, H} and the leftmost part of phase {L, H}. At these phases, the levels of both D and K are relatively higher than those at the other phases, which is consistent with experimental evidence [270]. At the late tumor stage (low ρ again), certain processes including epithelial-mesenchymal-transition (EMT) may trigger cancer metastasis, in which some small nests of cancer cells migrate to a remote location. From the new site, they initiate the development of a new metastasis with tumor cells that may have different identity and therefore require renewed immune recognition for tumor control.

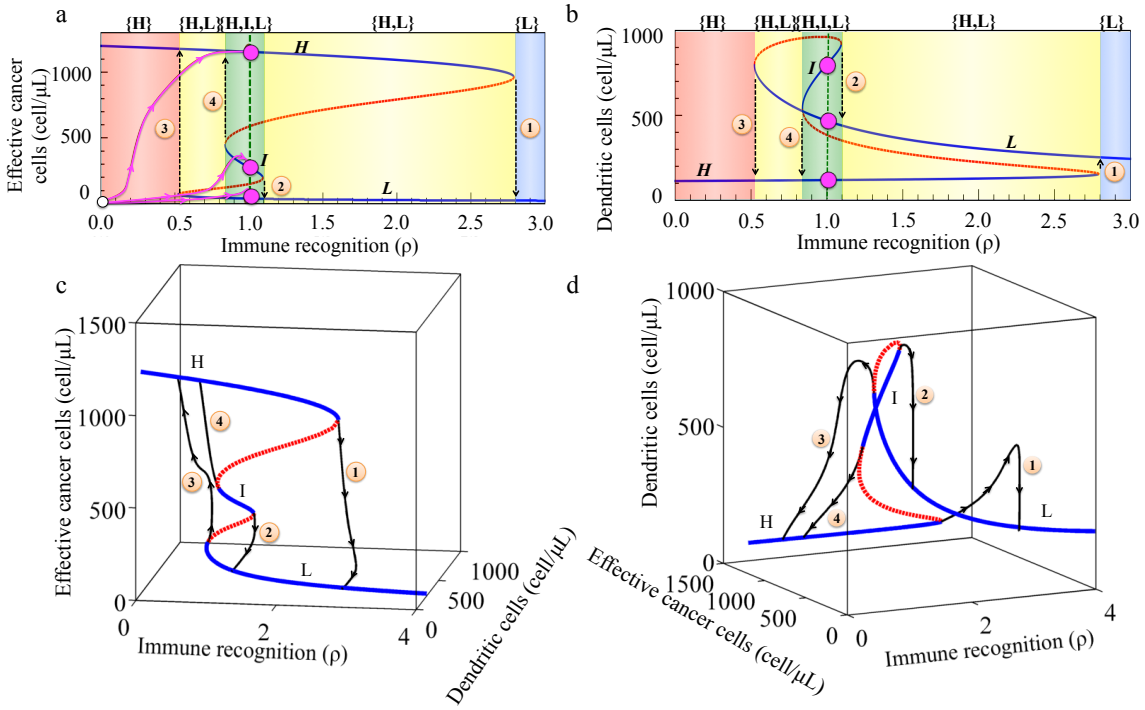


Figure 5.3 Bifurcation diagrams for the steady states as a function of the level of immune recognition ρ.

The diagrams show, for each value of immune recognition (ρ, x-axis), the steady state concentrations for the effective cancer cells (panel a) and the dendritic cells (panel b).

The states along the solid blue curves are the stable states, and those along the dashed red curves are the unstable states. The whole bifurcation curve is composed of segments of stable/unstable states, from which we define the high cancer state (H), the intermediate cancer state (I) and the low cancer state (L). Depending on the level of immune response, the system is at one of the four phases as represented by different background colors. Red is for the phase {H}, which has only the H state; yellow is for the phase {L,H}, which allows co-occurrence of both the L and H states, green is for the phase {L,I,H}, which allows occurrence of all three states; and blue is for the phase {L}, which has only the L state. During tumorigenesis when ρ changes, some transitions among different states could take place, as illustrated in dashed black lines and arrows, and are labeled with numbers. Three purple lines show the trajectories of development of tumor (with different rates of immune respond to tumor), and solid purple circles show the final states when full immune recognition is established. Panel c and d show the same bifurcation diagram in 3-D (with axes C, D and ρ) and in two perspectives.

Sensitivity to model parameters

We further explore sensitivity of the occurrence of the intermediate state and the tristability (Figure S4.3) to model parameters assuming that each tumor is characterized by a specific set of parameters. For this aim, we calculated the bifurcation diagram with respect to ρ for different sets of parameters. More specifically, we randomly vary all 36 model parameters simultaneously away from the original values by a uniform distribution with maximum of range of d% (chosen from 10%, 20% … 60%). We find that the intermediate (I) cancer state occurs for a significant range of model parameters (blue columns in Figure 5.4). For example, the (I) state occurs for about 50% of the parameter sets (or 50% chance) when the 36 parameters are varied randomly within a range of 40% about their original values. The chance for observing the tristability (red columns in Figure 5.4a) is slightly lower than that for the occurrence of the (I) state. Among those parameter sets in which the (I) state occurs (blue columns in Figure 5.4b) both the

average length of the (I) state and its standard deviation (Figure S4.4) go up with the range of the distribution from which the parameters are taken. Similar behavior is observed for the tristability (red columns in Figure 5.4b). For a larger perturbation of the parameters, the length of the (I) state is much wider than the length of the tristability (Figure 5.4b). The results imply that, when exosome exchange is operative, the intermediate cancer state and the tristability are characteristic features of the EM model for a wide range of model parameters.

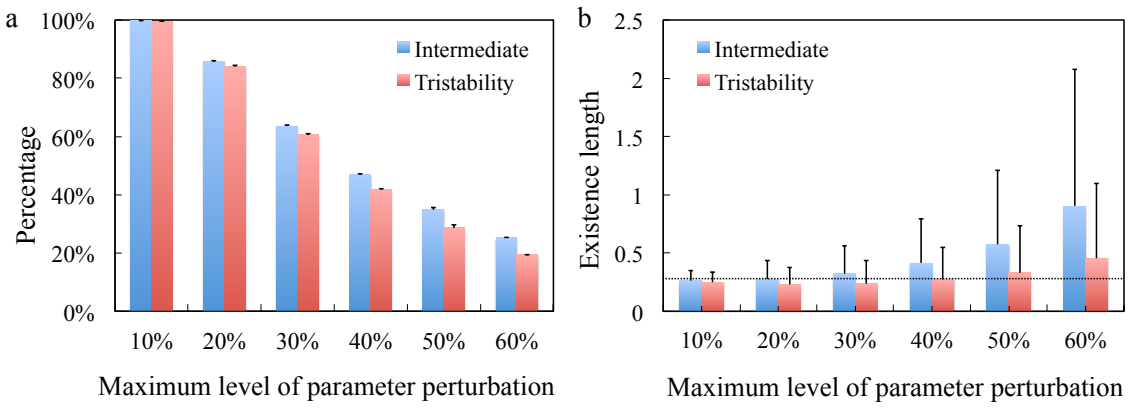


Figure 5.4 Sensitivity of the occurrence of the intermediate state to the model parameters.

In each test, all 36 parameters are randomly perturbed by a maximum of d% (chosen from 10%, 20% … 60%). Similar to Fig. 3, a bifurcation diagram with respect to the level of immune recognition ρ is generated to check the occurrence of the intermediate state and the co-occurrence of three stable states. Panel a shows the percentage of the occurrence of the intermediate state (blue columns) and the tristability (red columns) from a total of 10000 tests. From those cases in which the intermediate state or the tristability occurs, panel b shows the mean and standard deviation of the length of them in ρ. The dotted line shows the baseline, responding to the lengths of the intermediate state and the tristability for the case with the original (unperturbed) parameters (They overlap, as both lengths have the same values in ρ).

We further assess the effect of individual parameters on the occurrence of the (I) state and the tristability (Figure S4.5 and S4.6). We find that when any one of the parameters is varied by 10%, the length of the (I) state and of the tristability are only slightly affected (Figure S4.5a and S4.5b), but a variation of 30% in a single parameter can sometimes lead to significant difference (Figure S4.5d and S4.5d). The three most significant model parameters are associated with tumor growth (the Hill threshold of C self-activation C_C^0, the fold change of C self-activation λ_{CC}^+ and the basal proliferation rate of C, g_{C0} (Figure S4.7). These results imply that the occurrence of tristability and the (I) cancer state arises from a balance between cancer and immunity due to exosome exchange.

Building the cancer-immunity landscape

Motivated by the notion of Waddington landscape for cell differentiation [96], we proceed to introduce an effective landscape [97] corresponding to the dynamic states of the cancer-immunity interplay. Doing so helps to better understand the nature of the co-occurring cancer states and to design and assess possible therapeutic protocols by visualization of their effect as trajectories connecting different cancer states (or as a transition rate problem) in established landscape. In principle, the dynamics should be presented as a 4-dimension landscape whose axes are the densities of C, D, K, and the strength of the immune recognition ρ. Here, by assuming quasi-steady state for ρ, we consider 3-dimension landscapes (C, D, K) for different values of ρ. The landscape defines an effective potential computed as minus the logarithm of the probability (-

log(P)) of each dynamic state (C, D, K) in the presence of noise. The noise could originate from multiple sources, including cell-cell communications, birth-death processes of individual cells, coarse-graining of different cell types, and the effect of some other cancer/immune cell types or factors that are not included in the EM model.

In practice, it is easier to visualize the landscape in two dimensions. Therefore, we rely on projection of the 3-dimensional space onto a corresponding 2-dimensional one. For example, in Figure 5.2, the landscape is projected onto a 2-dimensional space whose axes are C and D. As shown in Figure S4.8 and Figure S4.9, we find that the landscape could be better presented if we replace D with a combination of D and K as one of the axes, since D and K are highly correlated (although in a nonlinear manner). We propose that the concept of the cancer-immunity landscape is very valuable for the design and assessment of effectiveness of novel therapeutic strategies. In particular, the landscape makes it clear that the desired/optimal treatments are not those that simply reduce the tumor size. As is illustrated further below, treatment effectiveness should be assessed by its efficacy to cross the dynamic barrier between high and low cancer states. It also shows that in most cases a direct transition from the H cancer state to the L state (H-to-L) is very hard to achieve. In the majority of cases, two-stage strategies are the desired ones – one treatment is used for effective H-to-I transition and a second treatment is used for effective I-to-L transition.

Effects of time delay on the cancer-immunity landscape

Communication between C, D and K cell types involves time delay. A longer time delay is expected for exosome exchange, because of the need for their production

and transport potentially across long distances. From a dynamic system perspective, time delay can have a significant effect on transitions between cancer states and on the effect of noise on the cancer-immunity interplay. We note that steady state solutions are not affected when the time delay is considered. However, the time delay might alter dynamics by changing the stability of the steady states (the cancer states) and the shape of the landscape. To this end, we evaluate the possible effects of time delays on the cancer-immunity interplay by performing simulations of the EM model (case I)

We examine the basin of attraction for each of the stable states (H, I and L) while incorporating the time delay in exosome exchange between C and D. From a dynamic systems perspective, the basin of attraction of each of the states is the domain of initial conditions from which the cancer/immunity evolve towards the particular dynamical state of the cancer-immune interplay (H, I or L). To construct the basin of attraction, we initially set the system at the high cancer (H) steady state. Assuming a subject is at the high cancer state (H), we simulate possible consequence of "hypothetical treatments". This is done by changing the population values of both C and D, while K cells are not affected. Such a treatment changes the system from the (H) state to a new state in the phase plane (Figure 5.5a). We then perform simulations with time delay (see Method section) for the dynamics starting from the newly induced location (new values of C and D). The simulations reveal that the dynamics after the initial induced deviations generate a trajectory that ends in one of the quasi-stable states (H, I and L). When starting from the (H) state, we define its corresponding basin of attraction as the region in the landscape (the phase plane), from which the dynamic trajectories evolve back towards the

(H) state. The basin of attraction for the (I) and (L) states can be constructed in a similar manner.

The upper left of Figure 5.5 shows the basins of attraction for the H (navy), I (blue) and L (light blue) states for case I and without time delay. An effective treatment should cause a transition from the (H) state to either the (I) or the (L) state. Our results indicate that, if a treatment aims to only reduce the C population, it requires a reduction of more than 90%, otherwise the system is still within the basin of attraction of the (H) state (see also left panel of Figure 5.5). Alternatively, it is more effective if the treatment targets the D population or both the C and D populations. By adding and increasing time delay, the basin of attraction for the (H) state increases linearly for small values of time delay and is saturated for large values of time delay (Figure 5.5), and that for the (L) state is mostly unaffected.

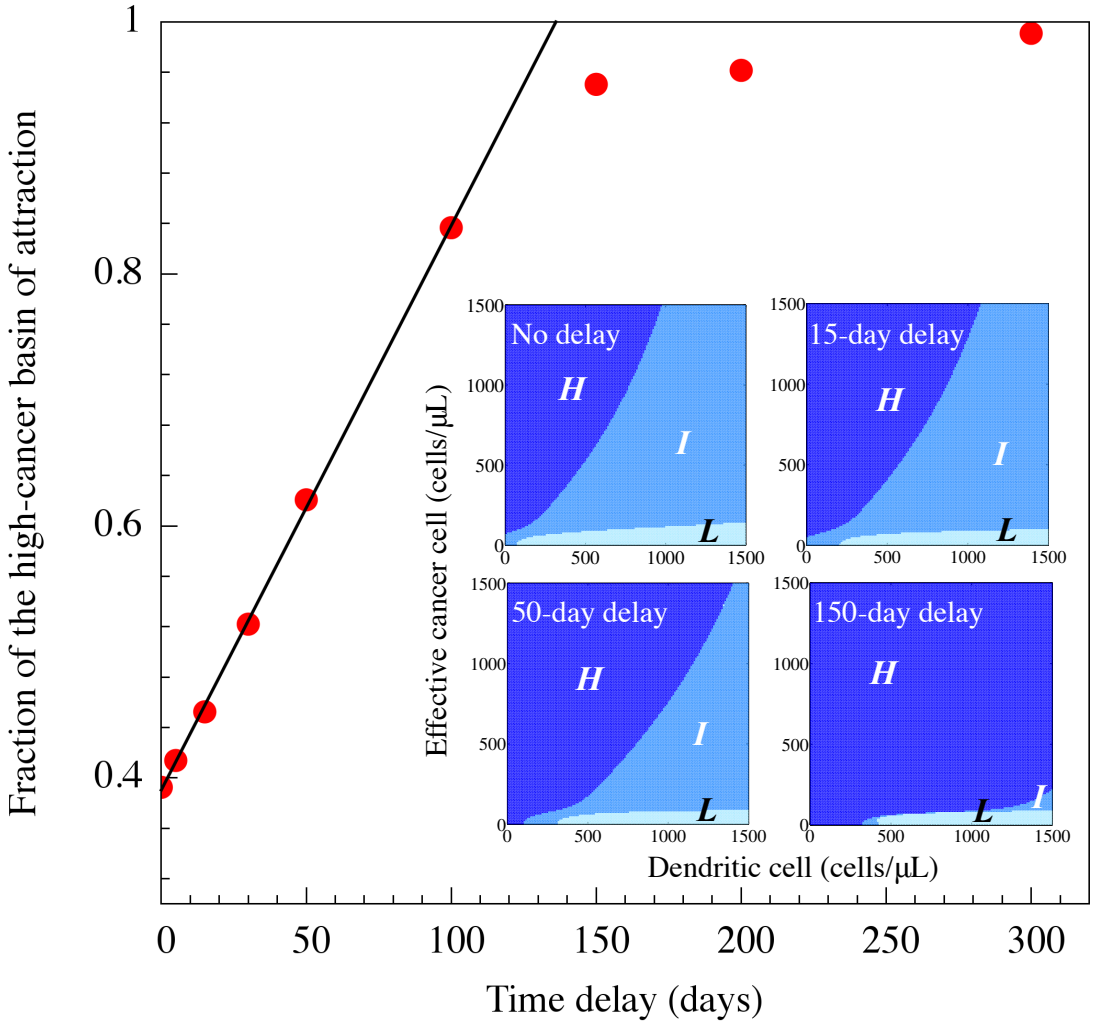


Figure 5.5 The effects of time delay in exosome-mediated communication on the cancer-immunity landscape.

The dynamics of the EM model are tested in multiple cases, each of which has different time delays for exosome-mediated communication between C and D (see Method section). In the test, the system is suddenly moved from the high cancer state (the initial state) to a new state in the phase plane as a consequence of "hypothetical treatment". Once released, the system follows its dynamic trajectory, and eventually reaches to one of the quasi-stable steady states. The basin of attraction for each quasi-stable steady state is defined as the region in the landscape (phase plane), from which the dynamic trajectories converge back to the target stable state. The whole phase planes is divided by the three basins for the high cancer state (navy), the intermediate state (blue), and the low cancer state (light blue). Panel a shows the phase planes for the cases of no delay (top

left), 15-day delay (top right), 50-day delay (bottom left) and 150-day delay (bottom right). The scatter plot in panel b shows the dependency of the fraction of the high-cancer basin of attraction (in area) on the value of the time delay.

Model-based design and assessment of therapeutic strategies

We present a design of several hypothetical therapeutic strategies and assessment of their effectiveness. Again, we assume that a subject's cancer is initially in the (H) state. Since the (H) state is located away from the L state in the phase-space (Figure 5.2b), it is hard to envision a treatment that can induce direct transitions from the (H) state to the (L) state (H2L). For this reason, in most cases (model parameter or cancer type) a more efficient protocol is likely to be based on a two-stage strategy: the first involves a treatment to induce the transition from the (H) state to the (I) state (H2I) and the second involves a treatment to induce the transition from the (I) state to the (L) state (I2L). We note that use of this more efficient two-stage strategy is suggested by the occurrence of the (I) state due to the exosome exchange.

To illustrate the concept in a quantitative manner, we mainly focus here on two types of common anti-cancer therapies – radiation therapy and DC based immunotherapy. The effect of the radiation therapy is incorporated into the EM model as elevation in apoptosis rates of C as well as of D and K during treatment. Note that apoptosis of D and K represents the inhibitory effect of radiation on the immune system [271,272]. DC immunotherapy is represented in the EM model as elevation in proliferation rates of D cells (see Method section).

The effect of hypothetical radiation therapy is shown in Figure 5.6. This example is composed of 30 one-per-day radiation sessions from day 10 (point 0) to day 40 (point 1) (Figure 5.6b). The population of D cells decreases during the 30 days (solid red line, Figure 5.6c). After the therapy ends, the population of D cells increases and reaches maximum at about day 90 (point 2), and then gradually decreases towards its initial level before therapy started (point 3, dashed back line). These results are in excellent agreement with clinical data for breast cancer patients after 30-day radiation therapy (Figure 5.6d, as adopted from ref. [272]). We note that the percentage of phagocytic activity of monocytes was measured in the clinical tests and not the density of D cells. However, the latter is correlated with the percentage of phagocytic activity of monocytes. We note that the model simulations are done independently of the clinical data, meaning that we do not fit the model parameters for this specific clinical test. For this reason, while in the model simulations the time recovery of the immune system to its value prior to the hypothetical radiation treatment is seven weeks, in the clinical test it is two weeks. For the treatment described in Figure 5.6, not only the immune system, but also the tumor load returns to its original high level as illustrated by the treatment trajectory in the cancer-immunity phase-space (Figure 5.6a). The trajectory exemplifies the fact that, although the radiation therapy reduces the tumor load (cancer level C) by 40%, after the treatment ends, C gradually increases (within a few weeks) back to its original (prior to the treatment) high level. Even if the radiation therapy is applied for 100 days instead of 30 days, so that C is further reduced by more than 50%, tumor load gradually returns to its original level (left panel of Figure 5.6a). It should be noted that a 100-day treatment is

not realistic since it is usually above the maximum permissible dose for the radiation therapy.

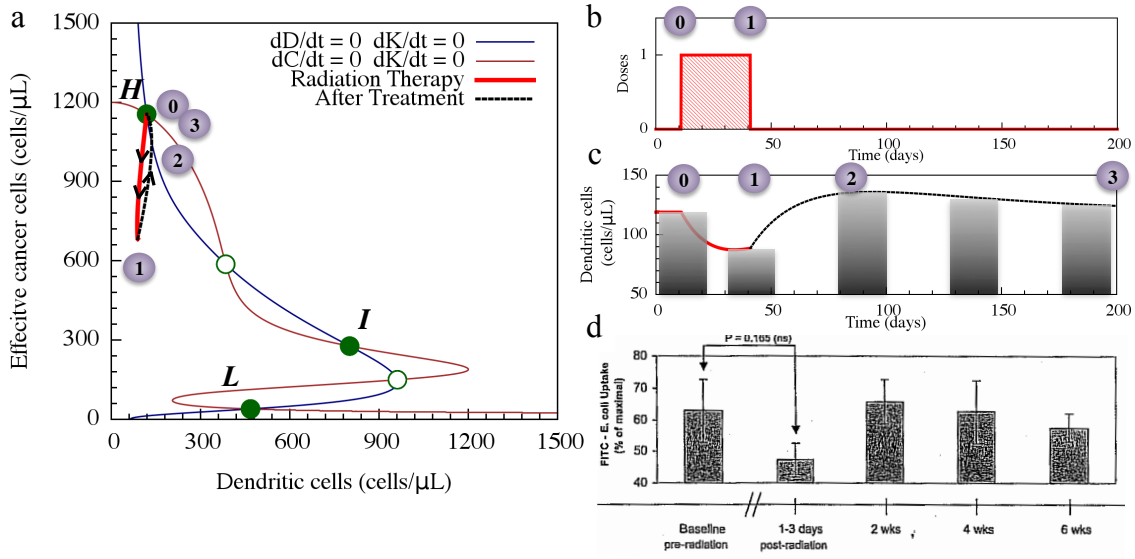


Figure 5.6 Assessment of hypothetical radiation therapy with the EM model.

The hypothetical therapy is composed of 30 one-per-day radiation sessions (modeled as an increase in the apoptosis rate of all cell types, see Method section) from day 10 (point 0) to day 40 (point 1, panel b). Panel a shows the dynamic trajectory in the phase plane, and panel c shows the time evolution of the concentrations of the dendritic cells. Panel d shows the experimental data on the phagocytic activity of monocytes for breast cancer patients after 30-day radiation therapy (data from [272]). The solid red lines and areas are the trajectory and doses for the radiation therapy, and the dashed black lines are the trajectory for the system after the treatment. Same as those in Figure 5.2, nullclines are shown in navy and brown solid lines, and steady states are shown as solid green circles (stable) and hollow green circles (unstable).

These above results suggest that radiation therapy alone may not be effective. As we illustrate next, a likely more effective strategy would be based on alternating combined radiation and immune therapy in a two-stage approach – an effective therapy

that can induce a transition from the high cancer state H to the intermediate state I (H2I stage), followed by an effective alternate therapy that can induce a transition from the intermediate cancer state I to the low cancer state L (I2L stage). Only in rare situations direct H2L transitions can be induced.

Assessment of simulated immunotherapy

The effect of hypothetical DC immunotherapy is shown in the right panel of Figure 5.7a. In this hypothetical therapy, the C level decreases to only about 85% of its original value at the end of therapy. Yet, since the therapy leads to a significant increase in the D level, it induces a transition from the high cancer state to the intermediate cancer state (H2I transition, right panel of Figure 5.7a). The induced transition is very slow though. Inspection of the time dependence of the H2I trajectory in the phase-plane reveals that it takes about three months for the system to approximate the unstable steady state (or the dynamic barrier) between the (H) and (I) states (hollow green circle in the phase-plan). Upon crossing this dynamic barrier, the system gradually approaches the intermediate state and converges to this state after about two years. Although the outcome of stand-alone immunotherapy is far better than that of the stand-alone radiation therapy, the slow convergence makes the process more sensitive to noise and increases the chance that random mutations or other events will cause the trajectory to change its course towards the high cancer state. In addition, application of too intense immunotherapy and/or application of immune boost for long periods of time increases the risk for side effects such as autoimmunity diseases, making stand-alone strong immunotherapy less desirable.

The Merit of alternating combined treatments

We proceed to test the efficiency of therapies that are based on alternating radiation therapy and immunotherapy. In the hypothetical example shown in the left panel of Figure 5.7b, we simulate three consecutive alternating treatments. Each treatment consists of combining a set of eight days of radiation therapy, followed by a set of eight days of immunotherapy. As illustrated in the left panel of Figure 5.7b, this hypothetical alternating combined therapy successfully induces H2I transition. Moreover, the alternating combined therapy uses far lower doses of radiation, and thus has significantly reduced side effects in comparison to a single stand-alone radiation therapy [273] and induces H2I transitions without the need to over boost of the immune system (in comparison to a stand-alone immune therapy) that also can have its own side effects.

Interestingly, inclusion of additional radiation therapy has an alarming effect – it causes a change in the therapy trajectory to converge back to the original high cancer state (right panel of Figure 5.7b). Moreover, even if the total dose of radiation therapy is reduced, it still can cause the same negative effect (Figure S4.10c). We also test treatments, in which both radiation therapy and immunotherapy are applied simultaneously on the same day, instead of consecutively on different days. In this case, the H2I transitions can be retained by using slightly reduced doses for radiation therapy and slightly milder amplification of the immune response (Figure S4.10a and S4.10b). However, the potential benefits might not justify the practical difficulties of applying both therapies simultaneously.

Therapeutic protocols for inducing I2L transitions

The dynamics of the cancer-immunity interplay in the vicinity of the intermediate (I) and low (L) cancer states are very different from those in the vicinity of the high (H) cancer state (Figure 5.2). For this reason, therapeutic protocols for inducing I2L transitions are expected to be different from those that are efficient for inducing H2I transitions. To devise an efficient protocol, we follow the same approach as before to assess the effectiveness of stand-alone radiation therapy, of stand-alone immunotherapy and of alternating combined therapy. As is illustrated in Figure S4.10d, stand-alone radiation therapy fails to induce the I2L transition similar to the inefficiency of radiation therapy to induce H2I transitions. However, stand-alone immunotherapy (five sets of ten days of immunotherapy, with ten-day break between any two consecutive sets) is found to be efficient to induce I2L transitions (left panel of Figure 5.7c). Surprisingly, alternating combined therapy turns out to be inefficient to induce the I2L (right panel of Figure 5.7c). The reason for the inefficiency is that the trajectory of the alternating combination is mostly orthogonal to the shortest path line that connects the (I) state with the dynamic barrier between the (I) and the (L) states (the unstable steady state marked by hollow green circle).

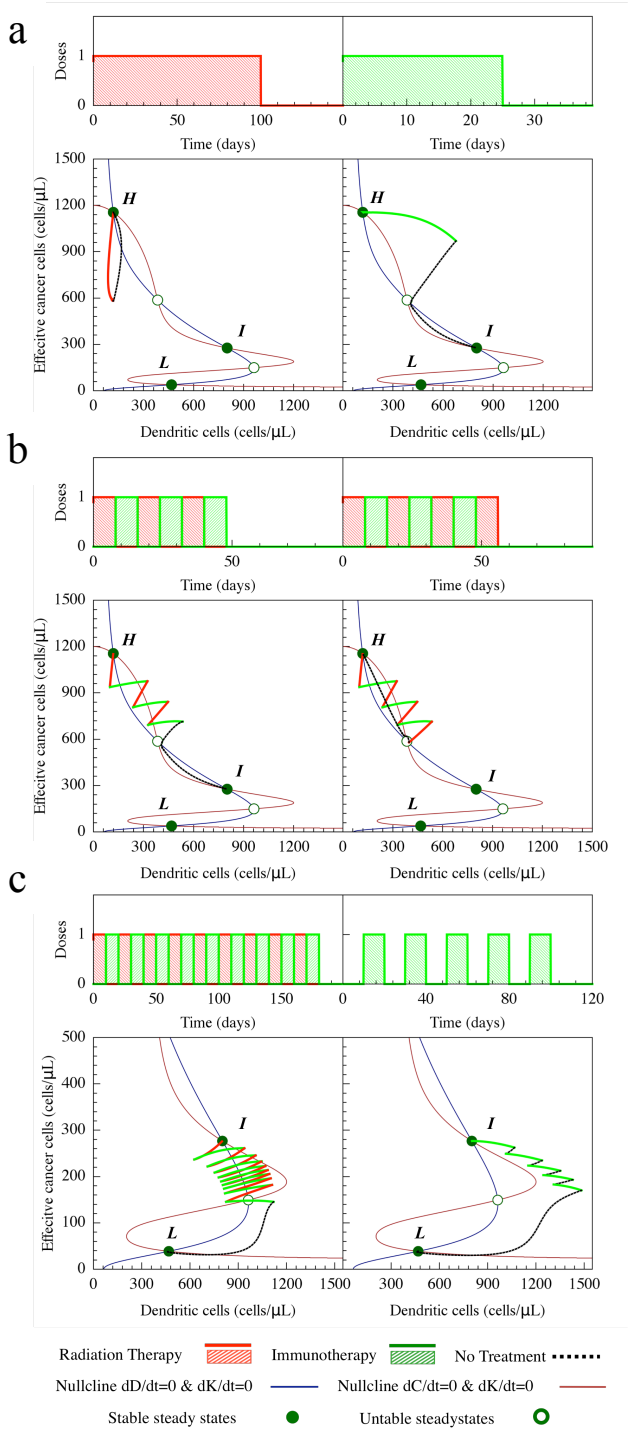


Figure 5.7 Assessment of various therapeutic strategies with the EM model.

Similar to Figure 5.6, we show in each case the dose usage (top part) and the dynamic trajectory in the phase plane (bottom part). The session of radiation therapy is shown in red and the session of DC immunotherapy (modeled as an increase in the proliferation rate of the dendritic cells, see Method section) is shown in green. Panel a shows the hypothetical treatment from the high cancer state (H) to the intermediate cancer state (I) in the case of the stand-alone radiation therapy (left), and the stand-alone DC immunotherapy (right). Panel b shows two cases of alternating combined therapy by both radiation therapy and immunotherapy. Panel c shows the hypothetical treatment from the intermediate cancer state (I) to the low cancer state (L) by the stand-alone immunotherapy (left), and an alternating combined treatment by both the radiation therapy and the immunotherapy (right).

Effect of time delay on treatment efficacy

To assess the effect of time delay on treatment efficacy, we convert the deterministic equations to include the time delay (see Method section). Specifically, we assess the case of 15-day delay for exosome exchange between C and D. Treatment simulations (Figure S4.11) show that treatment trajectories are similar to those for the cases without considering time delay. However, most treatments become less effective, because time delay makes transitions harder [274] as discussed in the proceeding time delay section.

The effect of noise on treatment efficacy

We also evaluate the effect of noise by turning the deterministic model into a stochastic model whose equations contain noise terms that represent white Gaussian noise (see Method section). The presence of large noise can induce spontaneous transitions between the different quasi-stable cancer states. To test the noise effect on the dynamics itself, we select the noise level to be sufficiently low to avoid noise induced

spontaneous transitions during the duration of treatments. Still, noise affects treatment effectiveness. Starting with the example presented previously, in which a 25-day immunotherapy in the absence of noise can trigger transition from the (H) to the (I) state (solid green line, Figure 5.8a), we find that the success rate for H2I transitions drops to 60% when the simulations start exactly from the levels of C, D and K of the (H) state. The reason is that it is possible for the noise to divert the treatment trajectory away from the path to cross the dynamic barrier between the (H) and the (I) states (Figure 5.8b, condition 3). However, in the presence of noise, the values of C, D and K vary in the vicinity of the exact steady state values before the treatment. Therefore the initial point at which the therapy starts can be close to but not exactly at the (H) state. Hence we continue to test the effects of the initial state on treatment outcome. To do so, we choose five different initial conditions 1-5 (insets of Figure 5.8a), whereby condition 3 is the exact (H) state, condition 1 is farthest from the transition state (the unstable steady state shown in hollow green circle) between the (H) and the (I) state, and condition 5 is closest to the dynamic barrier between the (H) and (I) states. As expected, the simulations reveal that the closer the initial state (when the therapy starts) to the dynamic barrier, the higher the probability to induce H2I transition, with the lowest probability being 20% when the therapy starts from condition 1. Notably, in the absence of noise, the therapy cannot induce H2I transition at all starting from this state. Considered together, the results indicate that noise sometimes can facilitate transitions during the treatment. In addition, when noise is included, there is low probability to induce direct H2L transitions that would not be possible in the absence of noise. Moreover, the results also demonstrate that it is crucial to start the treatment at the right time when the cancer-immunity

interplay is at a state as close as possible to the dynamic barrier for transitions to be induced. It follows that the cancer and immune states should be monitored closely and that therapy should be individualized and timed to increase the likelihood of the treatment success.

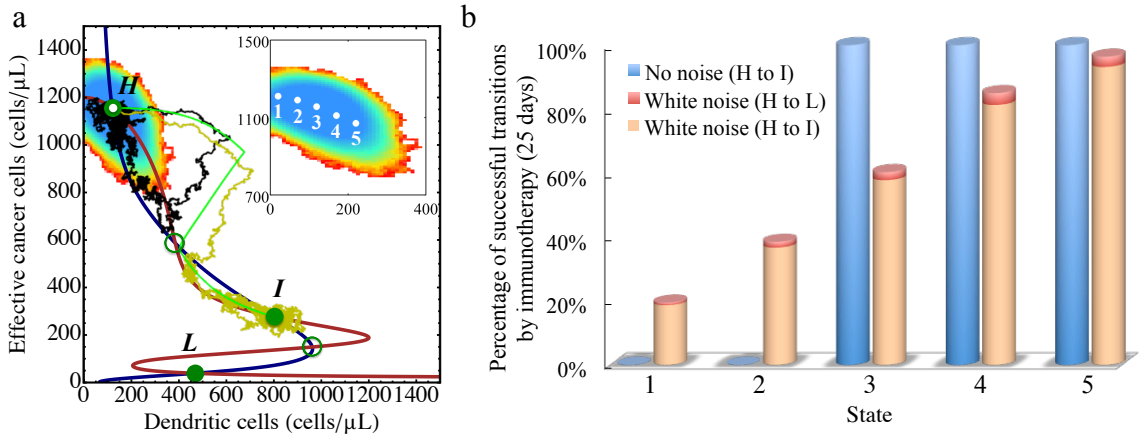


Figure 5.8 The effects of noise on the treatment by hypothetical immunotherapy.

Panel a shows the landscape for the system with white Gaussian noise. The noise level is not big enough to trigger transitions from high cancer state (H) to the intermediate cancer state (I). Without noise, a treatment of 25-day immunotherapy (solid green line, same as the right panel in Figure 5.7) causes the system to transit from H to I. With noise, the same treatment could either succeed (solid yellow line) or fail (solid black line) in making such a transition. We further select five different states (white dots in the inset) as the initial states for the treatment, where the state 3 is the original stable steady state. Panel B shows the percentage of successful transitions by the same treatment. In the absence of noise, the system always makes H to I transitions for the states 3, 4 and 5 (blue columns). In the presence of noise, the system has higher successful transitions from H to I when the initial state is closer to the saddle point (e.g. the state 5, orange columns). In some rare occasions (<5%), the system can also directly transit to the L state from the H state (red columns).

5.4. Discussion

In this study, we devised a theoretical framework (the EM model), to examine the contribution of exosome exchange to the intricate cancer-immunity interplay. We combined the complex interplay into a simplified generic model whose three components are cancer cells C, dendritic cells D (representing precursor, immature and mature types) and killer cells K (representing cytotoxic T cells, helper T cells, effector B cells and natural killer cells). Distinct from other models [249,251], the time evolution dynamics for each cell population was modeled with a set of nonlinear rate equations in a way similar to the analysis for gene regulatory circuits. Due to the specific cell-cell interactions among C, D and K cells, the EM model resembles the self-activating toggle switch, a gene circuit module that plays an important role in decision-making for many biological processes, e.g. epithelial-mesenchymal transitions [24,257], and cell differentiation [275–277]. It has been shown that such a circuit topology allows for three stable steady states [257]. For the current EM model, we also found that the cancer-immune system, at a certain level of immune recognition ρ and when exosome exchange is included, can allow three cancer states: high (H), intermediate (I) and low (L). The model prediction regarding the (I) state is consistent with the immunoediting theory [258]. Interestingly, owing to the special interactions among different cancer and immune cells, the immune system is most active in the intermediate state.

To study tumorigenesis, we incorporated the immune recognition represented by a variable ρ. During tumorigenesis, ρ gradually changes depending on the cancer stages. We proposed that, at the onset of a solid tumor, ρ starts from zero and increases to full

recognition ($\rho = 1$) at a certain response rate. We found that, depending on how fast the DCs recognize the new tumor, cancer development ends up at high, intermediate or low cancer states. Once the cancer population reaches a quasi-steady state, ρ gradually decreases due to immune escape. This dynamics of immune recognition are consistent with change in tumor immunogenicity in the immunoediting theory [258], However, it is still unclear if the equilibrium state exists, what derives and maintains the equilibrium and how to detect the equilibrium state [218].

We further analyzed the multi-stability of the model with respect to ρ within 2D and 3D bifurcation diagrams, from which we identified four different phases. In each of the phases, the model can be at one state, or stay in one of the multiple states. It is worth noting that the current EM model does not consider the situation in which the tumor actively escapes the immune suppression, which also contributes to reduced immune recognition. For example, the tumor can limit antigen presentation [278] during cancer treatment. We will take into account these effects by modeling the time evolution of ρ in a future study.

We also performed stochastic analysis on the EM model, from which we constructed a corresponding effective landscape. The effective landscape not only shows the relative stability of each steady state, but also provides an insight into cancer prognosis. Our findings imply that an effective treatment should not aim to simply reduce the size of the tumor, but to lead the dynamics of the cancer-immunity interplay to cross the dynamic barrier (effective potential barrier) during treatment from the (H) state to the (I) and from the (I) state to the (L) states. In many cases, effective therapeutic

protocols would be different for the H-to-I and the I-to-L transitions. Finally, we assessed the effects of time delays and biological noise on the design and effectiveness of therapeutic strategies. We note that noise effects of the cancer-immunity are still poorly understood [249].

The EM model enables us to evaluate the effectiveness of various treatment strategies. Our current model suggests that stand-alone radiation therapy is mostly ineffective, as the system usually goes back to the original quasi-stable state after the treatment. Radiation therapy may reduce not only the tumor load but also the immune response [271,272,279], making the system leans more towards the basin of the high cancer state. We also found that the immunotherapy is more effective. However, we may overestimate the effectiveness of the immunotherapy due to the following two factors. First, the slow convergence of cancer-immunity interplay to the L state during a stand-alone immunotherapy may increase the risk of failure due to some unpredicted factors de novo mutations in the tumor. Second, we do not know the maximum tolerated dosage for the immunotherapy, so it is possible that too much drug was applied causing intolerable side effects. For example, excessive use of immunotherapy may result in strong immune tolerance, and it can in turn reduce the immune recognition ρ or induce autoimmune diseases. Such potential effects will be taken into account in future studies.

While the EM model presented here turned out to be a valuable starting point, it calls for future studies in which the cell subpopulations in C, D and K are individually included in the EM model. For example, macrophages usually play a role in eliminating tumor cells and in priming Naïve T and B cells by antigen presentation, while tumor-

associated macrophages may stimulate tumor growth [223,280]. Moreover, regulatory T cells can induce immune tolerance, and may cause oscillatory cycles of the immune system [281]. Presumably, the effectiveness of cancer treatment could be affected by the dynamics of the regulatory T cells. It is crucial to investigate their roles in the population dynamics of the cancer-immune system.

References

1. Schlitt T, Palin K, Rung J, Dietmann S, Lappe M, Ukkonen E, et al. From Gene Networks to Gene Function. Genome Res. 2003;13: 2568–2576. doi:10.1101/gr.1111403

2. Han J-DJ. Understanding biological functions through molecular networks. Cell Res. 2008;18: 224–237. doi:10.1038/cr.2008.16

3. Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in Escherichia coli. Nature. 2000;403: 339–342. doi:10.1038/35002131

4. Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. Nature. 2000;403: 335–338. doi:10.1038/35002125

5. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet. 2003;33: 245–254. doi:10.1038/ng1089

6. Westerhoff HV, Palsson BO. The evolution of molecular biology into systems biology. Nat Biotech. 2004;22: 1249–1252. doi:10.1038/nbt1020

7. Alon U. An Introduction to Systems Biology: Design Principles of Biological Circuits. 1 edition. Boca Raton, FL: Chapman and Hall/CRC; 2006.

8. Seo S, Lewin HA. Reconstruction of metabolic pathways for the cattle genome. BMC Systems Biology. 2009;3: 33. doi:10.1186/1752-0509-3-33

9. Martin F-PJ, Dumas M-E, Wang Y, Legido‐Quigley C, Yap IKS, Tang H, et al. A top‐down systems biology view of microbiome‐mammalian metabolic interactions in a mouse model. Molecular Systems Biology. 2007;3: 112. doi:10.1038/msb4100153

10. Butcher EC, Berg EL, Kunkel EJ. Systems biology in drug discovery. Nat Biotech. 2004;22: 1253–1259. doi:10.1038/nbt1017

11. Lu M, Jolly MK, Levine H, Onuchic JN, Ben-Jacob E. MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. Proceedings of the National Academy of Sciences of the United States of America. 2013;110: 18174–9. doi:10.1073/pnas.1318192110

12. Zhang J, Tian X-J, Zhang H, Teng Y, Li R, Bai F, et al. TGF-β-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. Sci Signal. 2014;7: ra91. doi:10.1126/scisignal.2005304

13. Steinway SN, Zañudo JGT, Ding W, Rountree CB, Feith DJ, Loughran TP, et al. Network modeling of TGFβ signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint Sonic hedgehog and Wnt pathway activation. Cancer Res. 2014;74: 5963–5977. doi:10.1158/0008-5472.CAN-14-0225

14. Tyson JJ. Modeling the cell division cycle: cdc2 and cyclin interactions. PNAS. 1991;88: 7328–7332. doi:10.1073/pnas.88.16.7328

15. Li C, Wang J. Landscape and flux reveal a new global view and physical quantification of mammalian cell cycle. PNAS. 2014;111: 14130–14135. doi:10.1073/pnas.1408628111

16. Poirel CL, Rodrigues RR, Chen KC, Tyson JJ, Murali T m. Top-Down Network Analysis to Drive Bottom-Up Modeling of Physiological Processes. Journal of Computational Biology. 2013;20: 409–418. doi:10.1089/cmb.2012.0274

17. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28: 27–30.

18. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25: 25–29. doi:10.1038/75556

19. Lee W-P, Tzou W-S. Computational methods for discovering gene networks from expression data. Brief Bioinform. 2009;10: 408–423. doi:10.1093/bib/bbp028

20. Milo R, Phillips R. Cell Biology by the Numbers. 1 edition. New York, NY: Garland Science; 2015.

21. Strogatz SH. Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering. 1 edition. Cambridge, MA: Westview Press; 2001.

22. Heijer CD, Rheinboldt WC. On Steplength Algorithms for a Class of Continuation Methods. SIAM Journal on Numerical Analysis. 1981;18: 925–948.

23. Dhooge A, Govaerts W, Kuznetsov YA. MATCONT: A MATLAB Package for Numerical Bifurcation Analysis of ODEs. ACM Trans Math Softw. 2003;29: 141–164. doi:10.1145/779359.779362

24. Lu M, Jolly MK, Levine H, Onuchic JN, Ben-Jacob E. MicroRNA-based regulation of epithelial–hybrid–mesenchymal fate determination. PNAS. 2013;110: 18144–18149. doi:10.1073/pnas.1318192110

25. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. Cell. 2011;144: 646–674. doi:10.1016/j.cell.2011.02.013

26. Gupta GP, Massagué J. Cancer metastasis: building a framework. Cell. 2006;127: 679–695. doi:10.1016/j.cell.2006.11.001

27. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. J Clin Invest. 2009;119: 1420–1428. doi:10.1172/JCI39104

28. Jolly MK, Boareto M, Huang B, Jia D, Lu M, Ben-Jacob E, et al. Implications of the Hybrid Epithelial/Mesenchymal Phenotype in Metastasis. Front Oncol. 2015;5. doi:10.3389/fonc.2015.00155

29. Pankova K, Rosel D, Novotny M, Brabek J. The molecular mechanisms of transition between mesenchymal and amoeboid invasiveness in tumor cells. Cellular and molecular life sciences : CMLS. 2010;67: 63–71. doi:10.1007/s00018-009-0132-1

30. Ben-Jacob E, S. Coffey D, Levine H. Bacterial survival strategies suggest rethinking cancer cooperativity. Trends in Microbiology. 2012;20: 403–410. doi:10.1016/j.tim.2012.06.001

31. Bergert M, Chandradoss SD, Desai RA, Paluch E. Cell mechanics control rapid transitions between blebs and lamellipodia during migration. Proceedings of the National Academy of Sciences. 2012;109: 14434–14439.

32. Parri M, Taddei ML, Bianchini F, Calorini L, Chiarugi P. EphA2 Reexpression Prompts Invasion of Melanoma Cells Shifting from Mesenchymal to Amoeboid-like Motility Style. Cancer Research. 2009;69: 2072–2081. doi:10.1158/0008-5472.CAN-08-1845

33. Liu Y-J, Le Berre M, Lautenschlaeger F, Maiuri P, Callan-Jones A, Heuzé M, et al. Confinement and low adhesion induce fast amoeboid migration of slow mesenchymal cells. Cell. 2015;160: 659–672. doi:10.1016/j.cell.2015.01.007

34. Aung A, Seo YN, Lu S, Wang Y, Jamora C, del Álamo JC, et al. 3D traction stresses activate protease-dependent invasion of cancer cells. Biophys J. 2014;107: 2528–2537. doi:10.1016/j.bpj.2014.07.078

35. Sahai E, Marshall CJ. Differing modes of tumour cell invasion have distinct requirements for Rho/ROCK signalling and extracellular proteolysis. Nature. 2003;

36. Yamazaki D, Kurisu S, Takenawa T. Involvement of Rac and Rho signaling in cancer cell motility in 3D substrates. Oncogene. 2009;28: 1570–1583. doi:10.1038/onc.2009.2

37. Petrie RJ, Gavara N, Chadwick RS, Yamada KM. Nonpolarized signaling reveals two distinct modes of 3D cell migration. The Journal of Cell Biology. 2012;197: 439–455. doi:10.1083/jcb.201201124

38. Huang B, Lu M, Jolly MK, Tsarfaty I, Onuchic J, Ben-Jacob E. The three-way switch operation of Rac1/RhoA GTPase-based circuit controlling amoeboid-hybrid-mesenchymal transition. Scientific Reports. 2014;4. doi:10.1038/srep06449

39. Friedl P, Wolf K. Plasticity of cell migration: a multiscale tuning model. The Journal of cell biology. 2010;188: 11–9. doi:10.1083/jcb.200909003

40. Lu M, Jolly MK, Levine H, Onuchic JN, Ben-Jacob E. MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. Proceedings of the National Academy of Sciences. 2013;110: 18144–18149. doi:10.1073/pnas.1318192110

41. Guilluy C, Garcia-Mata R, Burridge K. Rho protein crosstalk: another social network? Trends Cell Biol. 2011;21: 718–726. doi:10.1016/j.tcb.2011.08.002

42. Boulter E, Garcia-Mata R, Guilluy C, Dubash A, Rossi G, Brennwald PJ, et al. Regulation of Rho GTPase crosstalk, degradation and activity by RhoGDI1. Nature Cell Biology. 2010;12: 477–483. doi:10.1038/ncb2049

43. Parri M, Chiarugi P. Rac and Rho GTPases in cancer cell motility control. Cell communication and signaling : CCS. 2010;8: 23. doi:10.1186/1478-811X-8-23

44. Sanz-Moreno V, Gadea G, Ahn J, Paterson H, Marra P, Pinner S, et al. Rac Activation and Inactivation Control Plasticity of Tumor Cell Movement. Cell. 2008;135: 510–523. doi:10.1016/j.cell.2008.09.043

45. Wyckoff JB, Pinner SE, Gschmeissner S, Condeelis JS, Sahai E. ROCK- and Myosin-Dependent Matrix Deformation Enables Protease-Independent Tumor-Cell Invasion In Vivo. Current Biology. 2006;16: 1515–1523. doi:10.1016/j.cub.2006.05.065

46. Kurisu S, Suetsugu S, Yamazaki D, Yamaguchi H, Takenawa T. Rac-WAVE2 signaling is involved in the invasive and metastatic phenotypes of murine melanoma cells. Oncogene. 2005;24: 1309–1319. doi:10.1038/sj.onc.1208177

47. Birchmeier C, Birchmeier W, Gherardi E, Vande Woude GF. Met, metastasis, motility and more. Nature reviews Molecular cell biology. 2003;4: 915–25. doi:10.1038/nrm1261

48. Graveel CR, DeGroot JD, Su Y, Koeman J, Dykema K, Leung S, et al. Met induces diverse mammary carcinomas in mice and is associated with human basal breast cancer. Proc Natl Acad Sci USA. 2009;106: 12909–12914. doi:10.1073/pnas.0810403106

49. Sierra JR, Tsao M-S. c-MET as a potential therapeutic target and biomarker in cancer. Ther Adv Med Oncol. 2011;3: S21–S35. doi:10.1177/1758834011422557

50. Zaritsky A, Natan2 S, Ben-Jacob E, Tsarfaty I. Emergence of HGF/SF-Induced Coordinated Cellular Motility. PLOS ONE. 2012;

51. Lamorte L, Royal I, Naujokas M, Park M. Crk adapter proteins promote an epithelial-mesenchymal-like transition and are required for HGF-mediated cell spreading and breakdown of epithelial adherens junctions. Mol Biol Cell. 2002;13: 1449–1461. doi:10.1091/mbc.01-10-0477

52. Laser-Azogui A, Diamant-Levi T, Israeli S, Roytman Y, Tsarfaty I. Met-induced membrane blebbing leads to amoeboid cell motility and invasion. Oncogene. 2013; doi:10.1038/onc.2013.138

53. Yamamura S, Saini S, Majid S, Hirata H, Ueno K, Chang I, et al. MicroRNA-34a suppresses malignant transformation by targeting c-Myc transcriptional complexes in human renal cell carcinoma. Carcinogenesis. 2012;33: 294–300. doi:10.1093/carcin/bgr286

54. Elson-Schwab I, Lorentzen A, Marshall CJ. MicroRNA-200 Family Members Differentially Regulate Morphological Plasticity and Mode of Melanoma Cell Invasion. PLoS ONE. 2010;5: e13176. doi:10.1371/journal.pone.0013176

55. Wang Z, Humphries B, Xiao H, Jiang Y, Yang C. MicroRNA-200b suppresses arsenic-transformed cell migration by targeting protein kinase Cα and Wnt5b-protein kinase Cα positive feedback loop and inhibiting Rac1 activation. J Biol Chem. 2014;289: 18373–18386. doi:10.1074/jbc.M114.554246

56. Humphries B, Wang Z, Oom AL, Fisher T, Tan D, Cui Y, et al. MicroRNA-200b targets protein kinase Cα and suppresses triple-negative breast cancer metastasis. Carcinogenesis. 2014;35: 2254–2263. doi:10.1093/carcin/bgu133

57. Kim D, Song J, Kim S, Park HM, Chun C-H, Sonn J, et al. MicroRNA-34a modulates cytoskeletal dynamics through regulating RhoA/Rac1 cross-talk in chondroblasts. J Biol Chem. 2012;287: 12501–12509. doi:10.1074/jbc.M111.264382

58. Ahn Y-H, Gibbons DL, Chakravarti D, Creighton CJ, Rizvi ZH, Adams HP, et al. ZEB1 drives prometastatic actin cytoskeletal remodeling by downregulating miR-34a expression. J Clin Invest. 2012;122: 3170–3183. doi:10.1172/JCI63608

59. Oberoi TK, Dogan T, Hocking JC, Scholz R-P, Mooz J, Anderson CL, et al. IAPs regulate the plasticity of cell migration by directly targeting Rac1 for degradation. The EMBO journal. 2011;31: 14–28.

60. Rolli-Derkinderen M. Phosphorylation of Serine 188 Protects RhoA from Ubiquitin/Proteasome-Mediated Degradation in Vascular Smooth Muscle Cells. Circulation Research. 2005;96: 1152–1160. doi:10.1161/01.RES.0000170084.88780.ea

61. Michaelson D, Silletti J, Murphy G, D'Eustachio P, Rush M, Philips MR. Differential localization of Rho GTPases in live cells regulation by hypervariable regions and RhoGDI binding. The Journal of cell biology. 2001;152: 111–126.

62. Fiegen D. Alternative Splicing of Rac1 Generates Rac1b, a Self-activating GTPase. Journal of Biological Chemistry. 2003;279: 4743–4749. doi:10.1074/jbc.M310281200

63. Zhang B, Zheng Y. Regulation of RhoA GTP hydrolysis by the GTPase-activating proteins p190, p50RhoGAP, Bcr, and 3BP-1. Biochemistry. 1998;37: 5249–5257.

64. Scheffzek K, Ahmadian MR. GTPase activating proteins: structural and functional insights 18 years after discovery. Cell Mol Life Sci. 2005;62: 3014–3038. doi:10.1007/s00018-005-5136-x

65. Gasmi-Seabrook GMC, Marshall CB, Cheung M, Kim B, Wang F, Jang YJ, et al. Real-time NMR Study of Guanine Nucleotide Exchange and Activation of RhoA by PDZ-RhoGEF. Journal of Biological Chemistry. 2009;285: 5137–5145. doi:10.1074/jbc.M109.064691

66. Wu X, Ramachandran S, Lin MJ, Cerione RA, Erickson JW. A Minimal Rac Activation Domain in the Unconventional Guanine Nucleotide Exchange Factor Dock180. Biochemistry. 2011;50: 1070–1080. doi:10.1021/bi100971y

67. Bourguignon LY, Zhu H, Shao L, Chen YW. Ankyrin-Tiam1 interaction promotes Rac1 signaling and metastatic breast tumor cell invasion and migration. J Cell Biol. 2000;150: 177–191.

68. Sako Y, Hibino K, Miyauchi T, Miyamoto Y, Ueda M, Yanagida T. Single-Molecule Imaging of Signaling Molecules in Living Cells. Single Molecules. 2000;1: 159–163. doi:10.1002/1438-5171(200006)1:2<159::AID-SIMO159>3.0.CO;2-4

69. Carlier M-F, Nioche P, Broutin-L'Hermite I, Boujemaa R, Le Clainche C, Egile C, et al. GRB2 Links Signaling to Actin Assembly by Enhancing Interaction of Neural Wiskott-Aldrich Syndrome Protein (N-WASp) with Actin-related Protein (ARP2/3) Complex. Journal of Biological Chemistry. 2000;275: 21946–21952. doi:10.1074/jbc.M000687200

70. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005;120: 15–20. doi:10.1016/j.cell.2004.12.035

71. Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009;19: 92–105. doi:10.1101/gr.082701.108

72. Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell. 2007;27: 91–105. doi:10.1016/j.molcel.2007.06.017

73. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP. Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nat Struct Mol Biol. 2011;18: 1139–1146. doi:10.1038/nsmb.2115

74. Yamazaki D, Kurisu S, Takenawa T. Regulation of cancer cell motility through actin reorganization. Cancer science. 2005;96: 379–86. doi:10.1111/j.1349-7006.2005.00062.x

75. Ridley AJ. Cell Migration: Integrating Signals from Front to Back. Science. 2003;302: 1704–1709. doi:10.1126/science.1092053

76. Lämmermann T, Bader BL, Monkley SJ, Worbs T, Wedlich-Söldner R, Hirsch K, et al. Rapid leukocyte migration by integrin-independent flowing and squeezing. Nature. 2008;453: 51–55. doi:10.1038/nature06887

77. Smith LA, Aranda-Espinoza H, Haun JB, Dembo M, Hammer DA. Neutrophil traction stresses are concentrated in the uropod during migration. Biophys J. 2007;92: L58–60. doi:10.1529/biophysj.106.102822

78. Yoshida K, Soldati T. Dissection of amoeboid movement into two mechanically distinct modes. J Cell Sci. 2006;119: 3833–3844. doi:10.1242/jcs.03152

79. Trinkaus JP. Surface activity and locomotion of Fundulus deep cells during blastula and gastrula stages. Developmental Biology. 1973;30: 68–103. doi:10.1016/0012-1606(73)90049-3

80. Diz-Muñoz A, Krieg M, Bergert M, Ibarlucea-Benitez I, Muller DJ, Paluch E, et al. Control of Directed Cell Migration In Vivo by Membrane-to-Cortex Attachment. PLOS Biology. 2010;8: e1000544. doi:10.1371/journal.pbio.1000544

81. Mishima T, Naotsuka M, Horita Y, Sato M, Ohashi K, Mizuno K. LIM-kinase is critical for the mesenchymal-to-amoeboid cell morphological transition in 3D matrices. Biochem Biophys Res Commun. 2010;392: 577–581. doi:10.1016/j.bbrc.2010.01.075

82. Leibowitz-Amit R, Tsarfaty G, Abargil Y, Yerushalmi GM, Horev J, Tsarfaty I. Mimp, a Mitochondrial Carrier Homologue, Inhibits Met-HGF/SF- Induced Scattering and Tumorigenicity by Altering Met-HGF/SF Signaling Pathways. Cancer Res. 2006;

83. Yerushalmi GM, Leibowitz-Amit R, Shaharabany M, Tsarfaty I. Met–HGF/SF Signal Transduction Induces Mimp, a Novel Mitochondrial Carrier Homologue, Which Leads to Mitochondrial Depolarization. Neoplasia. 2002;

84. Kaplan S, Bren A, Dekel E, Alon U. The incoherent feed-forward loop can generate non-monotonic input functions for genes. Molecular systems biology. 2008;4: 203. doi:10.1038/msb.2008.43

85. Palamidessi A, Frittoli E, Garré M, Faretta M, Mione M, Testa I, et al. Endocytic trafficking of Rac is required for the spatial restriction of signaling in cell migration. Cell. 2008;134: 135–147. doi:10.1016/j.cell.2008.05.034

86. Galliher-Beckley AJ, Schiemann WP. Grb2 binding to Tyr284 in T R-II is essential for mammary tumor growth and metastasis stimulated by TGF-. Carcinogenesis. 2008;29: 244–251. doi:10.1093/carcin/bgm245

87. Abella JV, Vaillancourt R, Frigault MM, Ponzo MG, Zuo D, Sangwan V, et al. The Gab1 scaffold regulates RTK-dependent dorsal ruffle formation through the adaptor Nck. Journal of Cell Science. 2010;123: 1306–1319. doi:10.1242/jcs.062570

88. Frank SA, Rosner MR. Nonheritable Cellular Variability Accelerates the Evolutionary Processes of Cancer. PLoS Biol. 2012;10: e1001296. doi:10.1371/journal.pbio.1001296

89. Balázsi G, van Oudenaarden A, Collins JJ. Cellular Decision Making and Biological Noise: From Microbes to Mammals. Cell. 2011;144: 910–925. doi:10.1016/j.cell.2011.01.030

90. Makrodouli E, Oikonomou E, Koc M, Andera L, Sasazuki T, Shirasawa S, et al. BRAF and RAS oncogenes regulate Rho GTPase pathways to mediate migration and invasion properties in human colon cancer cells: a comparative study. Mol Cancer. 2011;10: 118. doi:10.1186/1476-4598-10-118

91. Desai LP, Aryal AM, Ceacareanu B, Hassid A, Waters CM. RhoA and Rac1 are both required for efficient wound closure of airway epithelial cells. American Journal of Physiology - Lung Cellular and Molecular Physiology. 2004;287: L1134–L1144. doi:10.1152/ajplung.00022.2004

92. Geisbrecht ER, Montell DJ. A role for Drosophila IAP1-mediated caspase inhibition in Rac-dependent cell migration. Cell. 2004;118: 111–125. doi:10.1016/j.cell.2004.06.020

93. Park S-M, Gaur AB, Lengyel E, Peter ME. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. Genes Dev. 2008;22: 894–907. doi:10.1101/gad.1640608

94. Morizane R, Fujii S, Monkawa T, Hiratsuka K, Yamaguchi S, Homma K, et al. miR-34c attenuates epithelial-mesenchymal transition and kidney fibrosis with ureteral obstruction. Sci Rep. 2014;4. doi:10.1038/srep04578

95. Wolf K, Wu YI, Liu Y, Geiger J, Tam E, Overall C, et al. Multi-step pericellular proteolysis controls the transition from individual to collective cancer cell invasion. Nat Cell Biol. 2007;9: 893–904. doi:10.1038/ncb1616

96. Wang J, Zhang K, Xu L, Wang E. Quantifying the Waddington landscape and biological paths for development and differentiation. PNAS. 2011;108: 8257–8262. doi:10.1073/pnas.1017017108

97. Lu M, Onuchic J, Ben-Jacob E. Construction of an Effective Landscape for Multistate Genetic Switches. Phys Rev Lett. 2014;113: 078102. doi:10.1103/PhysRevLett.113.078102

98. Castellano E, Downward J. RAS Interaction with PI3K: More Than Just Another Effector Pathway. Genes & Cancer. 2011;2: 261–274. doi:10.1177/1947601911408079

99. Lambert JM, Lambert QT, Reuther GW, Malliri A, Siderovski DP, Sondek J, et al. Tiam1 mediates Ras activation of Rac by a PI(3)K-independent mechanism. Nature Cell Biology. 2002; doi:10.1038/ncb833

100. Janes PW, Daly RJ, deFazio A, Sutherland RL. Activation of the Ras signalling pathway in human breast cancer cells overexpressing erbB-2. Oncogene. 1994;9: 3601–3608.

101. Watanabe T, Tsuda M, Makino Y, Ichihara S, Sawa H, Minami A, et al. Adaptor molecule Crk is required for sustained phosphorylation of Grb2-associated binder 1 and hepatocyte growth factor-induced cell motility of human synovial sarcoma cell lines. Mol Cancer Res. 2006;4: 499–510. doi:10.1158/1541-7786.MCR-05-0141

102. Wang D, Paria BC, Zhang Q, Karpurapu M, Li Q, Gerthoffer WT, et al. A role for Gab1/SHP2 in thrombin activation of PAK1: gene transfer of kinase-dead PAK1 inhibits injury-induced restenosis. Circ Res. 2009;104: 1066–1075. doi:10.1161/CIRCRESAHA.109.196691

103. Schliekelman MJ, Taguchi A, Zhu J, Dai X, Rodriguez J, Celiktas M, et al. Molecular portraits of epithelial, mesenchymal, and hybrid States in lung adenocarcinoma and their relevance to survival. Cancer Res. 2015;75: 1789–1800. doi:10.1158/0008-5472.CAN-14-2535

104. Theveneau E, Mayor R. Cadherins in collective cell migration of mesenchymal cells. Current Opinion in Cell Biology. 2012;24: 677–684. doi:10.1016/j.ceb.2012.08.002

105. Palacios F, D'Souza-Schorey C. Modulation of Rac1 and ARF6 activation during epithelial cell scattering. J Biol Chem. 2003;278: 17395–17400. doi:10.1074/jbc.M300998200

106. Zhou JX, Huang S. Understanding gene circuits at cell-fate branch points for rational cell reprogramming. Trends Genet. 2011;27: 55–62. doi:10.1016/j.tig.2010.11.002

107. Jolly MK, Huang B, Lu M, Mani SA, Levine H, Ben-Jacob E. Towards elucidating the connection between epithelial–mesenchymal transitions and stemness. J R Soc Interface. 2014;11: 20140962. doi:10.1098/rsif.2014.0962

108. Jolly MK, Jia D, Boareto M, Mani SA, Pienta KJ, Ben-Jacob E, et al. Coupling the modules of EMT and stemness: A tunable "stemness window" model. Oncotarget. 2015;6: 25161–25174. doi:10.18632/oncotarget.4629

109. Jolly MK, Boareto M, Lu M, Onuchic JN, Clementi C, Ben-Jacob E. Operating principles of Notch–Delta–Jagged module of cell–cell communication. New Journal of Physics. 2015;17: 55021. doi:http://dx.doi.org/10.1088/1367-2630/17/5/055021

110. Boareto M, Jolly MK, Lu M, Onuchic JN, Clementi C, Ben-Jacob E. Jagged–Delta asymmetry in Notch signaling can give rise to a Sender/Receiver hybrid phenotype. Proceedings of the National Academy of Sciences. 2015;112: E402–E409. doi:10.1073/pnas.1416287112

111. Mierke CT. Physical view on migration modes. Cell Adh Migr. 2015;9: 367–379. doi:10.1080/19336918.2015.1066958

112. Hecht I, Bar-El Y, Balmer F, Natan S, Tsarfaty I, Schweitzer F, et al. Tumor Invasion Optimization by Mesenchymal-Amoeboid Heterogeneity. Scientific Reports. 2015;5: 10622. doi:10.1038/srep10622

113. Deakin NO, Turner CE. Distinct roles for paxillin and Hic-5 in regulating breast cancer cell morphology, invasion, and metastasis. Mol Biol Cell. 2011;22: 327–341. doi:10.1091/mbc.E10-09-0790

114. Grosse-Wilde A, d' Hérouël AF, McIntosh E, Ertaylan G, Skupin A, Kuestner RE, et al. Stemness of the hybrid Epithelial/Mesenchymal State in Breast Cancer and Its Association with Poor Survival. PLOS ONE. 2015;10: e0126522. doi:10.1371/journal.pone.0126522

115. Kircher M, Kelso J. High-throughput DNA sequencing--concepts and limitations. Bioessays. 2010;32: 524–536. doi:10.1002/bies.200900181

116. Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, et al. Yeast microarrays for genome wide parallel genetic and gene expression analysis. PNAS. 1997;94: 13057–13062.

117. Guo G, Huss M, Tong GQ, Wang C, Li Sun L, Clarke ND, et al. Resolution of Cell Fate Decisions Revealed by Single-Cell Gene Expression Analysis from Zygote to

Blastocyst. Developmental Cell. 2010;18: 675–685. doi:10.1016/j.devcel.2010.02.012

118. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. Nature. 2012;489: 101–108. doi:10.1038/nature11233

119. Moignard V, Macaulay IC, Swiers G, Buettner F, Schütte J, Calero-Nieto FJ, et al. Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. Nat Cell Biol. 2013;15: 363–372. doi:10.1038/ncb2709

120. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 2005;122: 947–956. doi:10.1016/j.cell.2005.08.020

121. Yan J, Wang H, Liu Y, Shao C. Analysis of Gene Regulatory Networks in the Mammalian Circadian Rhythm. PLOS Comput Biol. 2008;4: e1000193. doi:10.1371/journal.pcbi.1000193

122. Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. A circadian gene expression atlas in mammals: Implications for biology and medicine. PNAS. 2014;111: 16219–16224. doi:10.1073/pnas.1408886111

123. Smolen P, Baxter DA, Byrne JH. Mathematical Modeling of Gene Networks. Neuron. 2000;26: 567–580. doi:10.1016/S0896-6273(00)81194-0

124. Wang Z, Potoyan DA, Wolynes PG. Molecular stripping, targets and decoys as modulators of oscillations in the NF-κB/IκBα/DNA genetic network. Journal of The Royal Society Interface. 2016;13: 20160606. doi:10.1098/rsif.2016.0606

125. Narula J, Kuchina A, Lee DD, Fujita M, Süel GM, Igoshin OA. Chromosomal Arrangement of Phosphorelay Genes Couples Sporulation and DNA Replication. Cell. 2015;162: 328–337. doi:10.1016/j.cell.2015.06.012

126. Huang B, Lu M, Jolly MK, Tsarfaty I, Onuchic J, Ben-Jacob E. The three-way switch operation of Rac1/RhoA GTPase-based circuit controlling amoeboid-hybrid-mesenchymal transition. Sci Rep. 2014;4. doi:10.1038/srep06449

127. Lau K-Y, Ganguli S, Tang C. Function constrains network architecture and dynamics: A case study on the yeast cell cycle Boolean network. Phys Rev E. 2007;75: 051907. doi:10.1103/PhysRevE.75.051907

128. Zhao L, Sun T, Pei J, Ouyang Q. Mutation-induced protein interaction kinetics changes affect apoptotic network dynamic properties and facilitate oncogenesis. PNAS. 2015;112: E4046–E4054. doi:10.1073/pnas.1502126112

129. Li S, Zhu X, Liu B, Wang G, Ao P. Endogenous molecular network reveals two mechanisms of heterogeneity within gastric cancer. Oncotarget. 2015;6: 13607–13627.

130. Lei X, Tian W, Zhu H, Chen T, Ao P. Biological Sources of Intrinsic and Extrinsic Noise in cI Expression of Lysogenic Phage Lambda. Sci Rep. 2015;5. doi:10.1038/srep13597

131. Ge H, Qian H, Xie XS. Stochastic Phenotype Transition of a Single Cell in an Intermediate Region of Gene State Switching. Phys Rev Lett. 2015;114: 078101. doi:10.1103/PhysRevLett.114.078101

132. Kim JK, Josić K, Bennett MR. The relationship between stochastic and deterministic quasi-steady state approximations. BMC Systems Biology. 2015;9: 87. doi:10.1186/s12918-015-0218-3

133. Zhou JX, Samal A, d' Hérouël AF, Price ND, Huang S. Relative stability of network states in Boolean network models of gene regulation in development. Biosystems. 2016;142–143: 15–24. doi:10.1016/j.biosystems.2016.03.002

134. Liu J, Prindle A, Humphries J, Gabalda-Sagarra M, Asally M, Lee DD, et al. Metabolic co-dependence gives rise to collective oscillations within biofilms. Nature. 2015;523: 550–554. doi:10.1038/nature14660

135. Fei J, Singh D, Zhang Q, Park S, Balasubramanian D, Golding I, et al. Determination of in vivo target search kinetics of regulatory noncoding RNA. Science. 2015;347: 1371–1374. doi:10.1126/science.1258849

136. Meyer P, Cokelaer T, Chandran D, Kim KH, Loh P-R, Tucker G, et al. Network topology and parameter estimation: from experimental design methods to gene regulatory network kinetics using a community based approach. BMC Systems Biology. 2014;8: 13. doi:10.1186/1752-0509-8-13

137. Milo R, Jorgensen P, Moran U, Weber G, Springer M. BioNumbers—the database of key numbers in molecular and cell biology. Nucleic Acids Res. 2010;38: D750–D753. doi:10.1093/nar/gkp889

138. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network Motifs: Simple Building Blocks of Complex Networks. Science. 2002;298: 824–827. doi:10.1126/science.298.5594.824

139. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. Nature. 1999;402: C47–C52. doi:10.1038/35011540

140. Alon U. Network motifs: theory and experimental approaches. Nat Rev Genet. 2007;8: 450–461. doi:10.1038/nrg2102

141. Feng X, Hooshangi S, Chen D, Li G, Weiss R, Rabitz H. Optimizing Genetic Circuits by Global Sensitivity Analysis. Biophys J. 2004;87: 2195–2202. doi:10.1529/biophysj.104.044131

142. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally Sloppy Parameter Sensitivities in Systems Biology Models. PLOS Comput Biol. 2007;3: e189. doi:10.1371/journal.pcbi.0030189

143. Llamosi A, Gonzalez-Vargas AM, Versari C, Cinquemani E, Ferrari-Trecate G, Hersen P, et al. What Population Reveals about Individual Cell Identity: Single-Cell Parameter Estimation of Models of Gene Expression in Yeast. PLOS Comput Biol. 2016;12: e1004706. doi:10.1371/journal.pcbi.1004706

144. Meir E, von Dassow G, Munro E, Odell GM. Robustness, Flexibility, and the Role of Lateral Inhibition in the Neurogenic Network. Current Biology. 2002;12: 778–786. doi:10.1016/S0960-9822(02)00839-4

145. Prescott AM, McCollough FW, Eldreth BL, Binder BM, Abel SM. Analysis of Network Topologies Underlying Ethylene Growth Response Kinetics. Front Plant Sci. 2016; 1308. doi:10.3389/fpls.2016.01308

146. Li Z, Bianco S, Zhang Z, Tang C. Generic properties of random gene regulatory networks. Quant Biol. 2014;1: 253–260. doi:10.1007/s40484-014-0026-6

147. Ma W, Trusina A, El-Samad H, Lim WA, Tang C. Defining Network Topologies that Can Achieve Biochemical Adaptation. Cell. 2009;138: 760–773. doi:10.1016/j.cell.2009.06.013

148. Lu M, Jolly MK, Gomoto R, Huang B, Onuchic J, Ben-Jacob E. Tristability in Cancer-Associated MicroRNA-TF Chimera Toggle Switch. The Journal of Physical Chemistry B. 2013;117: 13164–13174. doi:10.1021/jp403156m

149. Wang J, Zhang K, Xu L, Wang E. Quantifying the Waddington landscape and biological paths for development and differentiation. PNAS. 2011;108: 8257–8262. doi:10.1073/pnas.1017017108

150. Huang S. Hybrid T-Helper Cells: Stabilizing the Moderate Center in a Polarized System. PLOS Biol. 2013;11: e1001632. doi:10.1371/journal.pbio.1001632

151. Sha W, Moore J, Chen K, Lassaletta AD, Yi C-S, Tyson JJ, et al. Hysteresis drives cell-cycle transitions in Xenopus laevis egg extracts. PNAS. 2003;100: 975–980. doi:10.1073/pnas.0235349100

152. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. PNAS. 1998;95: 14863–14868.

153. Saldanha AJ. Java Treeview--extensible visualization of microarray data. Bioinformatics. 2004;20: 3246–3248. doi:10.1093/bioinformatics/bth349

154. Huang S, Guo Y-P, May G, Enver T. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. Dev Biol. 2007;305: 695–713. doi:10.1016/j.ydbio.2007.02.036

155. Thiery JP, Acloque H, Huang RYJ, Nieto MA. Epithelial-Mesenchymal Transitions in Development and Disease. Cell. 2009;139: 871–890. doi:10.1016/j.cell.2009.11.007

156. Nieto MA, Huang RY-J, Jackson RA, Thiery JP. EMT: 2016. Cell. 2016;166: 21–45. doi:10.1016/j.cell.2016.06.028

157. Steinway SN, Zañudo JGT, Ding W, Rountree CB, Feith DJ, Loughran TP, et al. Network modeling of TGFβ signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint Sonic hedgehog and Wnt pathway activation. Cancer Res. 2014;74: 5963–5977. doi:10.1158/0008-5472.CAN-14-0225

158. Paranjape AN, Soundararajan R, Werden SJ, Joseph R, Taube JH, Liu H, et al. Inhibition of FOXC2 restores epithelial phenotype and drug sensitivity in prostate cancer cells with stem-cell properties. Oncogene. 2016; doi:10.1038/onc.2015.498

159. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. Nat Rev Mol Cell Biol. 2014;15: 178–196. doi:10.1038/nrm3758

160. Steinway SN, Zañudo JGT, Michel PJ, Feith DJ, Loughran TP, Albert R. Combinatorial interventions inhibit TGFβ-driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. npj Systems Biology and Applications. 2015;1: 15014. doi:10.1038/npjsba.2015.14

161. Cohen DPA, Martignetti L, Robine S, Barillot E, Zinovyev A, Calzone L. Mathematical Modelling of Molecular Pathways Enabling Tumour Cell Invasion and Migration. PLOS Comput Biol. 2015;11: e1004571. doi:10.1371/journal.pcbi.1004571

162. Reinhold WC, Sunshine M, Liu H, Varma S, Kohn KW, Morris J, et al. CellMiner: A Web-Based Suite of Genomic and Pharmacologic Tools to Explore Transcript and Drug Patterns in the NCI-60 Cell Line Set. Cancer Res. 2012;72: 3499–3511. doi:10.1158/0008-5472.CAN-12-1370

163. Albert I, Thakar J, Li S, Zhang R, Albert R. Boolean network simulations for life scientists. Source Code for Biology and Medicine. 2008;3: 16. doi:10.1186/1751-0473-3-16

164. Nishikawa S-I, Jakt LM, Era T. Embryonic stem-cell culture as a tool for developmental cell biology. Nat Rev Mol Cell Biol. 2007;8: 502–507. doi:10.1038/nrm2189

165. Cherry ABC, Daley GQ. Reprogramming cellular identity for regenerative medicine. Cell. 2012;148: 1110–1122. doi:10.1016/j.cell.2012.02.031

166. Loebel DAF, Watson CM, De Young RA, Tam PPL. Lineage choice and differentiation in mouse embryos and embryonic stem cells. Developmental Biology. 2003;264: 1–14. doi:10.1016/S0012-1606(03)00390-7

167. Stadtfeld M, Hochedlinger K. Induced pluripotency: history, mechanisms, and applications. Genes Dev. 2010;24: 2239–2263. doi:10.1101/gad.1963910

168. Daley GQ, Scadden DT. Prospects for Stem Cell-Based Therapy. Cell. 2008;132: 544–548. doi:10.1016/j.cell.2008.02.009

169. Chambers I, Silva J, Colby D, Nichols J, Nijmeijer B, Robertson M, et al. Nanog safeguards pluripotency and mediates germline development. Nature. 2007;450: 1230–1234. doi:10.1038/nature06403

170. Hayashi K, Lopes SMC de S, Tang F, Surani MA. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell Stem Cell. 2008;3: 391–401. doi:10.1016/j.stem.2008.07.027

171. Kumar RM, Cahan P, Shalek AK, Satija R, Jay DaleyKeyser A, Li H, et al. Deconstructing transcriptional heterogeneity in pluripotent stem cells. Nature. 2014;516: 56–61. doi:10.1038/nature13920

172. Singer ZS, Yong J, Tischler J, Hackett JA, Altinok A, Surani MA, et al. Dynamic Heterogeneity and DNA Methylation in Embryonic Stem Cells. Molecular Cell. 2014;55: 319–331. doi:10.1016/j.molcel.2014.06.029

173. Guo G, Huss M, Tong GQ, Wang C, Li Sun L, Clarke ND, et al. Resolution of Cell Fate Decisions Revealed by Single-Cell Gene Expression Analysis from Zygote to Blastocyst. Developmental Cell. 2010;18: 675–685. doi:10.1016/j.devcel.2010.02.012

174. Rao CV, Wolf DM, Arkin AP. Control, exploitation and tolerance of intracellular noise. Nature. 2002;420: 231–237. doi:10.1038/nature01258

175. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. Nature reviews Genetics. 2005;6: 451–64. doi:10.1038/nrg1615

176. Kircher M, Kelso J. High-throughput DNA sequencing--concepts and limitations. Bioessays. 2010;32: 524–536. doi:10.1002/bies.200900181

177. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10: 57–63. doi:10.1038/nrg2484

178. Herberg M, Roeder I. Computational modelling of embryonic stem-cell fate control. Development. 2015;142: 2250–2260. doi:10.1242/dev.116343

179. Chickarmane V, Peterson C. A Computational Model for Understanding Stem Cell, Trophectoderm and Endoderm Lineage Determination. PLOS ONE. 2008;3: e3478. doi:10.1371/journal.pone.0003478

180. Kalmar T, Lim C, Hayward P, Muñoz-Descalzo S, Nichols J, Garcia-Ojalvo J, et al. Regulated Fluctuations in Nanog Expression Mediate Cell Fate Decisions in Embryonic Stem Cells. PLOS Biology. 2009;7: e1000149. doi:10.1371/journal.pbio.1000149

181. Glauche I, Herberg M, Roeder I. Nanog Variability and Pluripotency Regulation of Embryonic Stem Cells - Insights from a Mathematical Model Analysis. PLOS ONE. 2010;5: e11238. doi:10.1371/journal.pone.0011238

182. Herberg M, Kalkan T, Glauche I, Smith A, Roeder I. A Model-Based Analysis of Culture-Dependent Phenotypes of mESCs. PLOS ONE. 2014;9: e92496. doi:10.1371/journal.pone.0092496

183. Chang R, Shoemaker R, Wang W. Systematic Search for Recipes to Generate Induced Pluripotent Stem Cells. PLOS Computational Biology. 2011;7: e1002300. doi:10.1371/journal.pcbi.1002300

184. Li C, Wang J. Quantifying Cell Fate Decisions for Differentiation and Reprogramming of a Human Stem Cell Network: Landscape and Biological Paths. Sarkar CA, editor. PLoS Computational Biology. 2013;9: e1003165. doi:10.1371/journal.pcbi.1003165

185. Huang B, Lu M, Jia D, Ben-Jacob E, Levine H, Onuchic J. Interrogating the Topological Robustness of Gene Regulatory Circuits. bioRxiv. 2016; 084962. doi:10.1101/084962

186. Zhang B, Wolynes PG. Stem cell differentiation as a many-body problem. PNAS. 2014;111: 10185–10190. doi:10.1073/pnas.1408561111

187. Sasai M, Kawabata Y, Makishi K, Itoh K, Terada TP. Time Scales in Epigenetic Dynamics and Phenotypic Heterogeneity of Embryonic Stem Cells. PLoS Comput Biol. 2013;9: e1003380. doi:10.1371/journal.pcbi.1003380

188. Duda RO, Hart PE, Stork DG. Pattern Classification. 2nd ed. New York: Wiley-Interscience; 2000.

189. Allen B, Kon M, Bar-Yam Y. A new phylogenetic diversity measure generalizing the shannon index and its application to phyllostomid bats. Am Nat. 2009;174: 236–243. doi:10.1086/600101

190. Marikawa Y, Alarcón VB. Establishment of trophectoderm and inner cell mass lineages in the mouse embryo. Mol Reprod Dev. 2009;76: 1019–1032. doi:10.1002/mrd.21057

191. Silva J, Nichols J, Theunissen TW, Guo G, van Oosten AL, Barrandon O, et al. Nanog Is the Gateway to the Pluripotent Ground State. Cell. 2009;138: 722–737. doi:10.1016/j.cell.2009.07.039

192. Schrode N, Saiz N, Di Talia S, Hadjantonakis A-K. GATA6 Levels Modulate Primitive Endoderm Cell Fate Choice and Timing in the Mouse Blastocyst. Developmental Cell. 2014;29: 454–467. doi:10.1016/j.devcel.2014.04.011

193. Ohnishi Y, Huber W, Tsumura A, Kang M, Xenopoulos P, Kurimoto K, et al. Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. Nat Cell Biol. 2014;16: 27–37. doi:10.1038/ncb2881

194. Wamaitha SE, del Valle I, Cho LTY, Wei Y, Fogarty NME, Blakeley P, et al. Gata6 potently initiates reprograming of pluripotent and differentiated cells to extraembryonic endoderm stem cells. Genes Dev. 2015;29: 1239–1255. doi:10.1101/gad.257071.114

195. Thomson M, Liu SJ, Zou L-N, Smith Z, Meissner A, Ramanathan S. Pluripotency Factors in Embryonic Stem Cells Regulate Differentiation into Germ Layers. Cell. 2011;145: 875–889. doi:10.1016/j.cell.2011.05.017

196. Hay DC, Sutherland L, Clark J, Burdon T. Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. Stem cells. 2004;22: 225–235.

197. Chen L, Yabuuchi A, Eminli S, Takeuchi A, Lu C-W, Hochedlinger K, et al. Cross-regulation of the Nanog and Cdx2 promoters. Cell Res. 2009;19: 1052–1061. doi:10.1038/cr.2009.79

198. Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K, Yagi R, et al. Interaction between Oct3/4 and Cdx2 Determines Trophectoderm Differentiation. Cell. 2005;123: 917–929. doi:10.1016/j.cell.2005.08.040

199. Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes & development. 2003;17: 126–140.

200. Strumpf D, Mao C-A, Yamanaka Y, Ralston A, Chawengsaksophak K, Beck F, et al. Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development. 2005;132: 2093–2102. doi:10.1242/dev.01801

201. Wang H, Dey SK. Roadmap to embryo implantation: clues from mouse models. Nat Rev Genet. 2006;7: 185–199. doi:10.1038/nrg1808

202. Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K, Yagi R, et al. Interaction between Oct3/4 and Cdx2 Determines Trophectoderm Differentiation. Cell. 2005;123: 917–929. doi:10.1016/j.cell.2005.08.040

203. Schrode N, Saiz N, Di Talia S, Hadjantonakis A-K. GATA6 Levels Modulate Primitive Endoderm Cell Fate Choice and Timing in the Mouse Blastocyst. Developmental Cell. 2014;29: 454–467. doi:10.1016/j.devcel.2014.04.011

204. Kumar RM, Cahan P, Shalek AK, Satija R, Jay DaleyKeyser A, Li H, et al. Deconstructing transcriptional heterogeneity in pluripotent stem cells. Nature. 2014;516: 56–61. doi:10.1038/nature13920

205. Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nature genetics. 2000;24: 372–376.

206. Waddington CH. The strategy of the genes: a discussion of some aspects of theoretical biology. Allen & Unwin; 1957.

207. Loh KM, Lim B. A Precarious Balance: Pluripotency Factors as Lineage Specifiers. Cell Stem Cell. 2011;8: 363–369. doi:10.1016/j.stem.2011.03.013

208. Morgani SM, Canham MA, Nichols J, Sharov AA, Migueles RP, Ko MSH, et al. Totipotent Embryonic Stem Cells Arise in Ground-State Culture Conditions. Cell Rep. 2013;3: 1945–1957. doi:10.1016/j.celrep.2013.04.034

209. Piotrowska-Nitsche K, Perea-Gomez A, Haraguchi S, Zernicka-Goetz M. Four-cell stage mouse blastomeres have different developmental properties. Development. 2005;132: 479–490. doi:10.1242/dev.01602

210. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature. 2008;453: 544–547. doi:10.1038/nature06965

211. Schütte J, Wang H, Antoniou S, Jarratt A, Wilson NK, Riepsaame J, et al. An experimentally validated network of nine haematopoietic transcription factors reveals mechanisms of cell state stability. Elife. 2016;5: e11469.

212. Lu M, Ma J. The Role of Shape in Determining Molecular Motions. Biophys J. 2005;89: 2395–2401. doi:10.1529/biophysj.105.065904

213. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: A synthesis. Proteins. 1995;21: 167–195. doi:10.1002/prot.340210302

214. Schlom J. Therapeutic Cancer Vaccines: Current Status and Moving Forward. JNCI J Natl Cancer Inst. 2012;104: 599–613. doi:10.1093/jnci/djs033

215. Leach DR, Krummel MF, Allison JP. Enhancement of antitumor immunity by CTLA-4 blockade. Science. 1996;271: 1734–1736.

216. Mougiakakos D, Choudhury A, Lladser A, Kiessling R, Johansson CC. Chapter 3 - Regulatory T Cells in Cancer. In: George F. Vande Woude and George Klein, editor. Advances in Cancer Research. Academic Press; 2010. pp. 57–117.

217. Sica A, Schioppa T, Mantovani A, Allavena P. Tumour-associated macrophages are a distinct M2 polarised population promoting tumour progression: Potential targets of anti-cancer therapy. European Journal of Cancer. 2006;42: 717–727. doi:10.1016/j.ejca.2006.01.003

218. Bhatia A, Kumar Y. Cancer-Immune Equilibrium: Questions Unanswered. Cancer Microenviron. 2011;4: 209–217. doi:10.1007/s12307-011-0065-8

219. Ben-Jacob E, Coffey DS, Levine H. Bacterial survival strategies suggest rethinking cancer cooperativity. Trends Microbiol. 2012;20: 403–410. doi:10.1016/j.tim.2012.06.001

220. Vesely MD, Kershaw MH, Schreiber RD, Smyth MJ. Natural Innate and Adaptive Immunity to Cancer. Annual Review of Immunology. 2011;29: 235–271. doi:10.1146/annurev-immunol-031210-101324

221. Igney FH, Krammer PH. Immune escape of tumors: apoptosis resistance and tumor counterattack. J Leukoc Biol. 2002;71: 907–920.

222. Hölzel M, Bovier A, Tüting T. Plasticity of tumour and immune cells: a source of heterogeneity and a cause for therapy resistance? Nat Rev Cancer. 2013;13: 365–376. doi:10.1038/nrc3498

223. Klimp AH, de Vries EGE, Scherphof GL, Daemen T. A potential role of macrophage activation in the treatment of cancer. Crit Rev Oncol Hematol. 2002;44: 143–161.

224. Smyth MJ, Thia KYT, Street SEA, Cretney E, Trapani JA, Taniguchi M, et al. Differential Tumor Surveillance by Natural Killer (Nk) and Nkt Cells. J Exp Med. 2000;191: 661–668. doi:10.1084/jem.191.4.661

225. Nagorsen D, Scheibenbogen C, Marincola FM, Letsch A, Keilholz U. Natural T Cell Immunity against Cancer. Clin Cancer Res. 2003;9: 4296–4303.

226. Toes REM, Ossendorp F, Offringa R, Melief CJM. CD4 T Cells and Their Role in Antitumor Immune Responses. J Exp Med. 1999;189: 753–756. doi:10.1084/jem.189.5.753

227. Kuby J. Immunology. New York: W H Freeman & Co; 1997.

228. Banchereau J, Steinman RM. Dendritic cells and the control of immunity. Nature. 1998;392: 245–252. doi:10.1038/32588

229. Silva RB da, Münz C. Natural killer cell activation by dendritic cells: balancing inhibitory and activating signals. Cell Mol Life Sci. 2011;68: 3505–3518. doi:10.1007/s00018-011-0801-8

230. Schroder K, Hertzog PJ, Ravasi T, Hume DA. Interferon-γ: an overview of signals, mechanisms and functions. J Leukoc Biol. 2004;75: 163–189. doi:10.1189/jlb.0603252

231. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer. 2012;12: 323–334. doi:10.1038/nrc3261

232. Mueller MM, Fusenig NE. Friends or foes — bipolar effects of the tumour stroma in cancer. Nat Rev Cancer. 2004;4: 839–849. doi:10.1038/nrc1477

233. Holmes EC. Immunology of tumor infiltrating lymphocytes. Ann Surg. 1985;201: 158–163.

234. Harper J, Sainson RCA. Regulation of the anti-tumour immune response by cancer-associated fibroblasts. Seminars in Cancer Biology. 2014;25: 69–77. doi:10.1016/j.semcancer.2013.12.005

235. Vaupel P, Kallinowski F, Okunieff P. Blood Flow, Oxygen and Nutrient Supply, and Metabolic Microenvironment of Human Tumors: A Review. Cancer Res. 1989;49: 6449–6465.

236. Vaupel P, Kelleher DK, Höckel M. Oxygen status of malignant tumors: pathogenesis of hypoxia and significance for tumor therapy. Semin Oncol. 2001;28: 29–35.

237. Munich S, Sobo-Vujanovic A, Buchser WJ, Beer-Stolz D, Vujanovic NL. Dendritic cell exosomes directly kill tumor cells and activate natural killer cells via

TNF superfamily ligands. OncoImmunology. 2012;1: 1074–1083. doi:10.4161/onci.20897

238. Marcus ME, Leonard JN. FedExosomes: Engineering Therapeutic Biological Nanoparticles that Truly Deliver. Pharmaceuticals (Basel). 2013;6: 659–680. doi:10.3390/ph6050659

239. Février B, Raposo G. Exosomes: endosomal-derived vesicles shipping extracellular messages. Current Opinion in Cell Biology. 2004;16: 415–421. doi:10.1016/j.ceb.2004.06.003

240. Camussi G, Deregibus M-C, Bruno S, Grange C, Fonsato V, Tetta C. Exosome/microvesicle-mediated epigenetic reprogramming of cells. Am J Cancer Res. 2010;1: 98–110.

241. Zhang H-G, Grizzle WE. Exosomes and Cancer: A Newly Described Pathway of Immune Suppression. Clinical Cancer Research. 2011;17: 959–964. doi:10.1158/1078-0432.CCR-10-1489

242. Peinado H, Alečković M, Lavotshkin S, Matei I, Costa-Silva B, Moreno-Bueno G, et al. Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. Nature Medicine. 2012;18: 883–891. doi:10.1038/nm.2753

243. Viaud S, Terme M, Flament C, Taieb J, André F, Novault S, et al. Dendritic Cell-Derived Exosomes Promote Natural Killer Cell Activation and Proliferation: A Role for NKG2D Ligands and IL-15Rα. PLoS ONE. 2009;4: e4942. doi:10.1371/journal.pone.0004942

244. Matsumoto K, Morisaki T, Kuroki H, Kubo M, Onishi H, Nakamura K, et al. Exosomes secreted from monocyte-derived dendritic cells support in vitro naive CD4+ T cell survival through NF-κB activation. Cellular Immunology. 2004;231: 20–29. doi:10.1016/j.cellimm.2004.11.002

245. Hao S, Bai O, Li F, Yuan J, Laferte S, Xiang J. Mature dendritic cells pulsed with exosomes stimulate efficient cytotoxic T-lymphocyte responses and antitumour immunity. Immunology. 2007;120: 90–102. doi:10.1111/j.1365-2567.2006.02483.x

246. Yu S, Liu C, Su K, Wang J, Liu Y, Zhang L, et al. Tumor exosomes inhibit differentiation of bone marrow dendritic cells. The Journal of Immunology. 2007;178: 6867–6875.

247. Liu C, Yu S, Zinn K, Wang J, Zhang L, Jia Y, et al. Murine Mammary Carcinoma Exosomes Promote Tumor Growth by Suppression of NK Cell Function. J Immunol. 2006;176: 1375–1385. doi:10.4049/jimmunol.176.3.1375

248. Banasiak J, Chaplain MAJ, Miękisz J. Multiscale Problems in the Life Sciences. Capasso V, Lachowicz M, editors. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008.

249. Eftimie R, Bramson JL, Earn DJD. Interactions Between the Immune System and Cancer: A Brief Review of Non-spatial Mathematical Models. Bulletin of Mathematical Biology. 2011;73: 2–32. doi:10.1007/s11538-010-9526-3

250. Minelli A, Topputo F, Bernelli-Zazzera F. Controlled Drug Delivery in Cancer Immunotherapy: Stability, Optimization, and Monte Carlo Analysis. SIAM Journal on Applied Mathematics. 2011;71: 2229–2245. doi:10.1137/100815190

251. Robertson-Tessi M, El-Kareh A, Goriely A. A mathematical model of tumor–immune interactions. Journal of Theoretical Biology. 2012;294: 56–73. doi:10.1016/j.jtbi.2011.10.027

252. Pappalardo F, Pennisi M, Ricupito A, Topputo F, Bellone M. Induction of T-cell memory by a dendritic cell vaccine: a computational model. Bioinformatics. 2014;30: 1884–1891. doi:10.1093/bioinformatics/btu059

253. Wilkie KP, Hahnfeldt P. Mathematical models of immune-induced cancer dormancy and the emergence of immune evasion. Interface Focus. 2013;3: 20130010–20130010. doi:10.1098/rsfs.2013.0010

254. Rihan FA, Abdel Rahman DH, Lakshmanan S, Alkhajeh AS. A time delay model of tumour–immune system interactions: Global dynamics, parameter estimation, sensitivity analysis. Applied Mathematics and Computation. 2014;232: 606–623. doi:10.1016/j.amc.2014.01.111

255. Mehrian M, Asemani D, Arabameri A, Pourgholaminejad A, Hadjati J. Modeling of tumor growth in dendritic cell-based immunotherapy using artificial neural networks. Computational Biology and Chemistry. 2014;48: 21–28. doi:10.1016/j.compbiolchem.2013.09.007

256. Palsson S, Hickling TP, Bradshaw-Pierce EL, Zager M, Jooss K, Peter JO, et al. The development of a fully-integrated immune response model (FIRM) simulator of the immune response through integration of multiple subset models. BMC systems biology. 2013;7: 95.

257. Lu M, Jolly MK, Gomoto R, Huang B, Onuchic J, Ben-Jacob E. Tristability in Cancer-Associated MicroRNA-TF Chimera Toggle Switch. J Phys Chem B. 2013;117: 13164–13174. doi:10.1021/jp403156m

258. Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. Nat Immunol. 2002;3: 991–998. doi:10.1038/ni1102-991

259. Qu J-L, Qu X-J, Zhao M-F, Teng Y-E, Zhang Y, Hou K-Z, et al. Gastric cancer exosomes promote tumour cell proliferation through PI3K/Akt and MAPK/ERK activation. Digestive and Liver Disease. 2009;41: 875–880. doi:10.1016/j.dld.2009.04.006

260. Gerosa F, Baldani-Guerra B, Nisii C, Marchesini V, Carra G, Trinchieri G. Reciprocal Activating Interaction between Natural Killer Cells and Dendritic Cells. J Exp Med. 2002;195: 327–333. doi:10.1084/jem.20010938

261. Vivier E, Raulet DH, Moretta A, Caligiuri MA, Zitvogel L, Lanier LL, et al. Innate or Adaptive Immunity? The Example of Natural Killer Cells. Science. 2011;331: 44–49. doi:10.1126/science.1198687

262. Nelson BH. CD20+ B Cells: The Other Tumor-Infiltrating Lymphocytes. J Immunol. 2010;185: 4977–4982. doi:10.4049/jimmunol.1001323

263. Clayton A, Mitchell JP, Court J, Mason MD, Tabi Z. Human Tumor-Derived Exosomes Selectively Impair Lymphocyte Responses to Interleukin-2. Cancer Res. 2007;67: 7458–7466. doi:10.1158/0008-5472.CAN-06-3456

264. Thomas DA, Massagué J. TGF-beta directly targets cytotoxic T cell functions during tumor evasion of immune surveillance. Cancer Cell. 2005;8: 369–380. doi:10.1016/j.ccr.2005.10.012

265. Hanke N, Alizadeh D, Katsanis E, Larmonier N. Dendritic cell tumor killing activity and its potential applications in cancer immunotherapy. Crit Rev Immunol. 2013;33: 1–21.

266. Wolfers J, Lozier A, Raposo G, Regnault A, Théry C, Masurier C, et al. Tumor-derived exosomes are a source of shared tumor rejection antigens for CTL cross-priming. Nat Med. 2001;7: 297–303. doi:10.1038/85438

267. Menetrier-Caux C, Montmain G, Dieu MC, Bain C, Favrot MC, Caux C, et al. Inhibition of the Differentiation of Dendritic Cells From CD34+ Progenitors by Tumor Cells: Role of Interleukin-6 and Macrophage Colony-Stimulating Factor. Blood. 1998;92: 4778–4791.

268. Shampine LF, Thompson S. Solving DDEs in Matlab. Applied Numerical Mathematics. 2001;37: 441–458. doi:10.1016/S0168-9274(00)00055-6

269. Kloeden PE, Platen E. Numerical Solution of Stochastic Differential Equations. Springer; 1992.

270. Scarlett UK, Rutkowski MR, Rauwerdink AM, Fields J, Escovar-Fadul X, Baird J, et al. Ovarian cancer progression is controlled by phenotypic changes in dendritic cells. J Exp Med. 2012;209: 495–506. doi:10.1084/jem.20111413

271. Formenti SC, Demaria S. Systemic effects of local radiotherapy. The Lancet Oncology. 2009;10: 718–726. doi:10.1016/S1470-2045(09)70082-8

272. Standish LJ, Torkelson C, Hamill FA, Yim D, Hill-Force A, Fitzpatrick A, et al. Immune defects in breast cancer patients after radiotherapy. Journal of the Society for Integrative Oncology. 2008;6: 110.

273. Levy A, Chargari C, Cheminant M, Simon N, Bourgier C, Deutsch E. Radiation therapy and immunotherapy: Implications for a combined cancer treatment. Critical Reviews in Oncology/Hematology. 2013;85: 278–287. doi:10.1016/j.critrevonc.2012.09.001

274. Gupta C, López JM, Ott W, Josić K, Bennett MR. Transcriptional Delay Stabilizes Bistable Gene Networks. Phys Rev Lett. 2013;111: 058104. doi:10.1103/PhysRevLett.111.058104

275. Li C, Wang J. Quantifying Cell Fate Decisions for Differentiation and Reprogramming of a Human Stem Cell Network: Landscape and Biological Paths. PLoS Comput Biol. 2013;9: e1003165. doi:10.1371/journal.pcbi.1003165

276. Zhang B, Wolynes PG. Stem cell differentiation as a many-body problem. PNAS. 2014;111: 10185–10190. doi:10.1073/pnas.1408561111

277. Huang S. Hybrid T-Helper Cells: Stabilizing the Moderate Center in a Polarized System. PLoS Biol. 2013;11: e1001632. doi:10.1371/journal.pbio.1001632

278. Maeurer MJ, Gollin SM, Martin D, Swaney W, Bryant J, Castelli C, et al. Tumor escape from immune recognition: lethal recurrent melanoma in a patient associated with downregulation of the peptide transporter protein TAP-1 and loss of expression of the immunodominant MART-1/Melan-A antigen. J Clin Invest. 1996;98: 1633–1641. doi:10.1172/JCI118958

279. Uh S, Lee SM, Kim HT, Chung Y, Kim YH, Park C, et al. THe effect of radiation therapy on immune function in patients with squamous cell lung carcinoma. Chest. 1994;105: 132–137. doi:10.1378/chest.105.1.132

280. Quatromoni JG, Eruslanov E. Tumor-associated macrophages: function, phenotype, and link to prognosis in human lung cancer. Am J Transl Res. 2012;4: 376–389.

281. Coventry BJ, Ashdown ML, Quinn MA, Markovic SN, Yatomi-Clarke SL, Robinson AP. CRP identifies homeostatic immune oscillations in cancer patients: a potential treatment targeting tool? Journal of Translational Medicine. 2009;7: 102. doi:10.1186/1479-5876-7-102

Supplementary Information

Appendix 1: Towards elucidating the plasticity of cell migration phenotypes during cancer metastasis

Table S1.1 - Experimental evidences for Rac1/RhoA regulatory circuits.

Regulations	Involved Molecules*	References	Cell lines
Rac1-GTP inhibits RhoA-GTP	p190RhoGAP	Anjaruwee et al., 2003[1] William et al., 2001[2] Che-Hung et al., 2008[3]	Breast cancer cells, Hela cells, Rat1 fibroblasts
RhoA-GTP inhibits Rac1-GTP	FilGAP	Koji et al., 2012[4]	Breast, lung, prostate and colorectal adenocarcinoma cells
	ARHGAP22	Victoria et al., 2008[5]	Melanoma cells
Rac1-GTP self-activation	IRSp53/EP58	Yosuke et al., 2004[6]	Breast cancer cells, Melanoma cells, Fibrosarcoma cells et al.
	Hem-1	Orion et al., 2006[7]	Human neutrophil-like cells
RhoA-GTP self-activation	Dia1	Thomas et al., 2007[8]	Breast cancer cells, Human embryonic kidney cancer cell
RhoA transcriptional self-activation	c-Myc	Chia-Hsin et al., 2010[9] Xose, 2010[10]	Breast cancer cells, Fibroblasts
Grb2 activates Rac1-GTP	Ras	Esther et al., 2011[11] John et al., 2002[12]	Breast cancer cells, Human embryonic kidney 293 cells et al.

		Peter et al., 1994[13]	
Gab1 activates Rac1-GTP	Crk	Takuya et al., 2006[14]	Breast cancer cells, Human synovial sarcoma cells
Gab1 activates RhoA-GTP	LARG	Dong et al., 2009[15]	Rat vascular smooth muscle cells

* Most regulations are mediated indirectly through GEFs or GAPs, such as p190RhoGAP.

Table S1.2 - Parameters for Rac1/RhoA regulatory Circuit

Parameters	Value	Unit	Description
Rac1			
g_{R_c}	3.4×105	molecules/h	Production rate
K_{R_c}	0.1	h^{-1}	Degradation rate for Rac1-GDP
$K_{R_c^*}$	0.1	h^{-1}	Degradation rate for Rac1-GTP
gdi_R_c	2.0×103	h^{-1}	Binding rate for GDI to Rac1-GDP
$dgdi\_R_c$	2.0×103	h^{-1}	Dissociation rate for Rac1-GDI
$gtp\_R_c i$	0.54	h^{-1}	Intrinsic GTP loading rate
$gtp\_R_c B$	19.46	h^{-1}	Activated GTP loading rate
$gtp\_R_c A$	530	h^{-1}	Activated GTP loading rate by self-activation
$gtp\_R_c A_0$	1.9×106	molecules	Threshold for self-activation
$n_{gtp\_R_c A}$	4	-	Hill coefficient for self-activation
$dgtp\_R_c i$	6.6	h^{-1}	Intrinsic GTP hydrolysis rate
$dgtp\_R_c B$	105.4	h^{-1}	Activated GTP hydrolysis rate
$dgtp\_R_c A$	198	h^{-1}	Activated GTP hydrolysis rate by RhoA
$dgtp\_R_c A_0$	1.2×106	molecules	Threshold for RhoA inactivation
$n_{dgtp\_R_c A}$	4	-	Hill coefficient for RhoA inactivation
RhoA			
g_{R_h}	1.6×105	molecules/h	Basal production rate
$g_{R_h A}$	3.4×105	molecules/h	Excitatory production rate
$g_{R_h A_0}$	8.0×105	molecules	Threshold for transcriptionally self-activation
$n_{g_{R_h A}}$	4	-	Hill coefficient for self-activation
K_{R_h}	0.1	h^{-1}	Degradation rate for RhoA-GDP
$K_{R_h^*}$	0.1	h^{-1}	Degradation rate for RhoA-GTP
gdi_R_h	2.0×103	h^{-1}	Binding rate for GDI to RhoA-GDP
$dgdi\_R_h$	2.0×103	h^{-1}	Dissociation rate for RhoA-GDI
$gtp\_R_h i$	0.54	h^{-1}	Intrinsic GTP loading rate
$gtp\_R_h B$	109.46	h^{-1}	Activated GTP loading rate
$gtp\_R_h A$	196	h^{-1}	Activated GTP loading rate by self-activation
$gtp\_R_h A_0$	1.0×106	molecules	Threshold for self-activation
$n_{gtp\_R_h A}$	4	-	Hill coefficient for self-activation
$dgtp\_R_h i$	1.32	h^{-1}	Intrinsic GTP hydrolysis rate
$dgtp\_R_h B$	308.68	h^{-1}	Activated GTP hydrolysis rate
$dgtp\_R_h A$	89	h^{-1}	Activated GTP hydrolysis rate by Rac1
$dgtp\_R_h A_0$	1.3×106	molecules	Threshold for Rac1 inactivation

$n_{dgtp\_R_hA}$	4	-	Hill coefficient for Rac1 inactivation
Signals			
$gtp\_R_cI_1$	240	h^{-1}	Activated GTP loading rate for Rac1 by Grb2
$gtp\_R_cI_{10}$	5.0×105	molecules	Threshold for Grb2 activation on Rac1
$n_{gtp\_R_cI_1}$	2	-	Hill coefficient for Grb2 activation on Rac1
$gtp\_R_hI_2$	240	h^{-1}	Activated GTP loading rate for RhoA by Gab1
$gtp\_R_hI_{20}$	5.0×105	molecules	Threshold for Gab1 activation on RhoA
$n_{gtp\_R_hI_2}$	2	-	Hill coefficient for Gab1 activation on RhoA
$gtp\_R_cI_2$	90	h^{-1}	Activated GTP loading rate for Rac1 by Gab1
$gtp\_R_cI_{20}$	5.0×105	molecules	Threshold for Gab1 activation on Rac1
$n_{gtp\_R_cI_2}$	2	-	Hill coefficient for Gab1 activation on Rac1

Table S1.3 - Parameters Table for the coupled gene regulatory circuit

Parameters	Value	Unit	Description
Rac1			
g_{R_c}	3.4×10^5	molecules/h	Production rate
K_{R_c}	0.1	h^{-1}	Degradation rate for Rac1-GDP
$K_{R_c^*}$	0.1	h^{-1}	Degradation rate for Rac1-GTP
gdi_R_c	2.0×10^3	h^{-1}	Binding rate for GDI to Rac1-GDP
$dgdi\_R_c$	2.0×10^3	h^{-1}	Dissociation rate for Rac1-GDI
gtp_R_ci	0.54	h^{-1}	Intrinsic GTP loading rate
gtp_R_cB	19.46	h^{-1}	Activated GTP loading rate
gtp_R_cA	530	h^{-1}	Activated GTP loading rate by self-activation
$gtp\_R_cA_0$	1.9×10^6	molecules	Threshold for self-activation
$n_{gtp\_R_cA}$	4	-	Hill coefficient for self-activation
$dgtp\_R_ci$	6.6	h^{-1}	Intrinsic GTP hydrolysis rate
$dgtp\_R_cB$	105.4	h^{-1}	Activated GTP hydrolysis rate
$dgtp\_R_cA$	198	h^{-1}	Activated GTP hydrolysis rate by RhoA
$dgtp\_R_cA_0$	1.2×10^6	molecules	Threshold for RhoA inactivation
$n_{dgtp\_R_cA}$	4	-	Hill coefficient for RhoA inactivation
RhoA			
g_{R_h}	1.6×10^5	molecules/h	Basal production rate
g_{R_hA}	3.4×10^5	molecules/h	Excitatory production rate
$g_{R_hA_0}$	8.0×10^5	molecules	Threshold for transcriptionally self-activation
$n_{g_{R_hA}}$	4	-	Hill coefficient for self-activation
K_{R_h}	0.1	h^{-1}	Degradation rate for RhoA-GDP
$K_{R_h^*}$	0.1	h^{-1}	Degradation rate for RhoA-GTP
gdi_R_h	2.0×10^3	h^{-1}	Binding rate for GDI to RhoA-GDP
$dgdi\_R_h$	2.0×10^3	h^{-1}	Dissociation rate for RhoA-GDI
gtp_R_hi	0.54	h^{-1}	Intrinsic GTP loading rate
gtp_R_hB	109.46	h^{-1}	Activated GTP loading rate
gtp_R_hA	196	h^{-1}	Activated GTP loading rate by self-activation
$gtp\_R_hA_0$	1.0×10^6	molecules	Threshold for self-activation
$n_{gtp\_R_hA}$	4	-	Hill coefficient for self-activation
$dgtp\_R_hi$	1.32	h^{-1}	Intrinsic GTP hydrolysis rate
$dgtp\_R_hB$	308.68	h^{-1}	Activated GTP hydrolysis rate
$dgtp\_R_hA$	89	h^{-1}	Activated GTP hydrolysis rate by Rac1
$dgtp\_R_hA_0$	1.3×10^6	molecules	Threshold for Rac1 inactivation

$n_{dgtp\_R_hA}$	4	-	Hill coefficient for Rac1 inactivation
c-Met Signals			
$gtp\_R_cI_1$	240	h^{-1}	Activated GTP loading rate for Rac1 by Grb2
$gtp\_R_cI_{10}$ *	5.0×10^5	molecules	Threshold for Grb2 activation on Rac1
$n_{gtp\_R_cI_1}$	2	-	Hill coefficient for Grb2 activation on Rac1
$gtp\_R_hI_2$	240	h^{-1}	Activated GTP loading rate for RhoA by Gab1
$gtp\_R_hI_{20}$ *	5.0×10^5	molecules	Threshold for Gab1 activation on RhoA
$n_{gtp\_R_hI_2}$	2	-	Hill coefficient for Gab1 activation on RhoA
$gtp\_R_cI_2$	90	h^{-1}	Activated GTP loading rate for Rac1 by Gab1
$gtp\_R_cI_{20}$	5.0×10^5	molecules	Threshold for Gab1 activation on Rac1
$n_{gtp\_R_cI_2}$	2	-	Hill coefficient for Gab1 activation on Rac1
microRNAs			
n_μ	2	-	Hill coefficient for microRNAs
l_0	1	-	Coefficient for $L(\mu)$
l_1	0.6	-	Coefficient for $L(\mu)$
l_2	0.3	-	Coefficient for $L(\mu)$
γ_{m1}	0.04	h^{-1}	Coefficient for $Y_m(\mu)$
γ_{m2}	0.2	h^{-1}	Coefficient for $Y_m(\mu)$
K_{mR_c}	0.5	h^{-1}	Degradation rate for mRNA of Rac1
K_{mR_c}	0.5	h^{-1}	Degradation rate for mRNA of RhoA
μ_0	1.0×10^4	molecules	Threshold for microRNA regulation

* There are two thresholds for Grb2 and Gab1 regulations are changed to 2.5×10^5 molecules for Figure. 2.13.

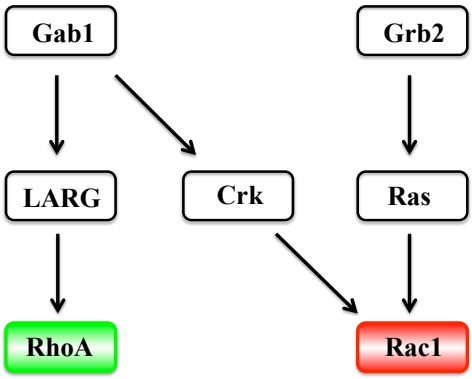
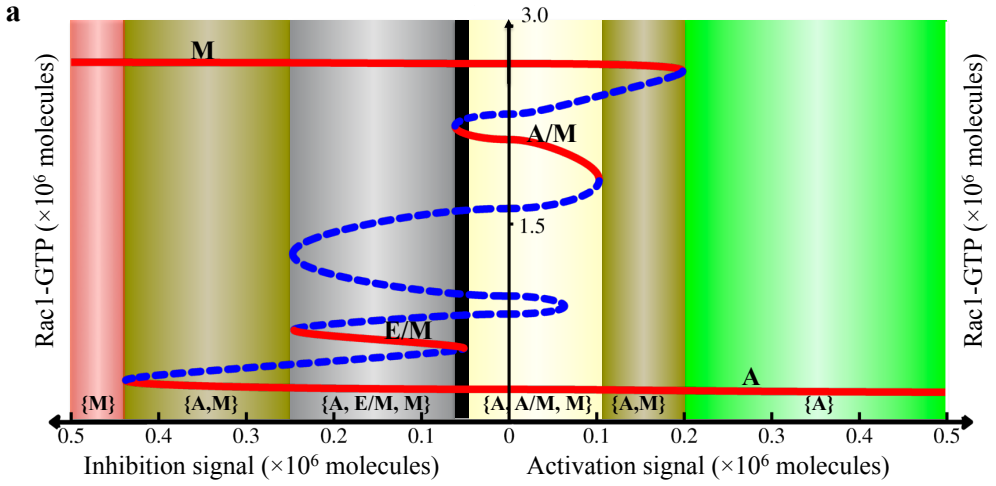
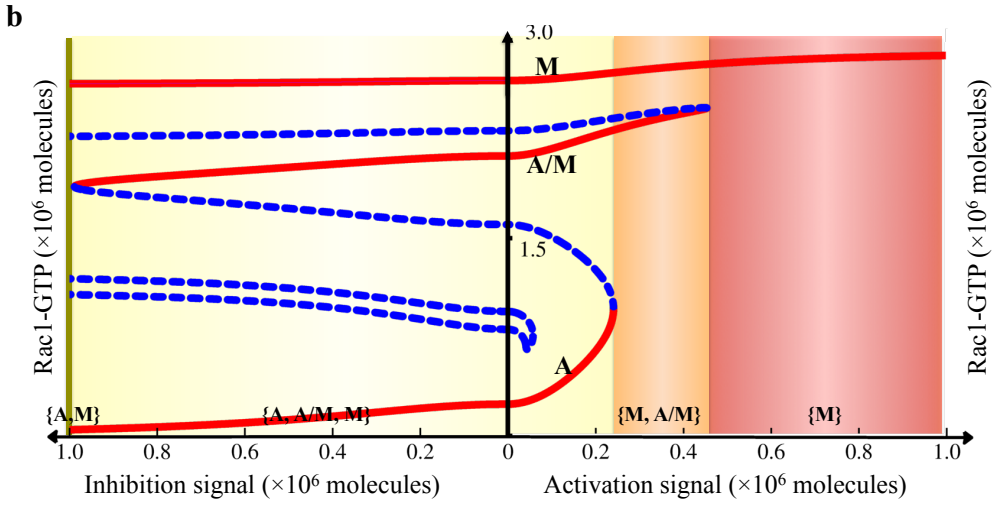
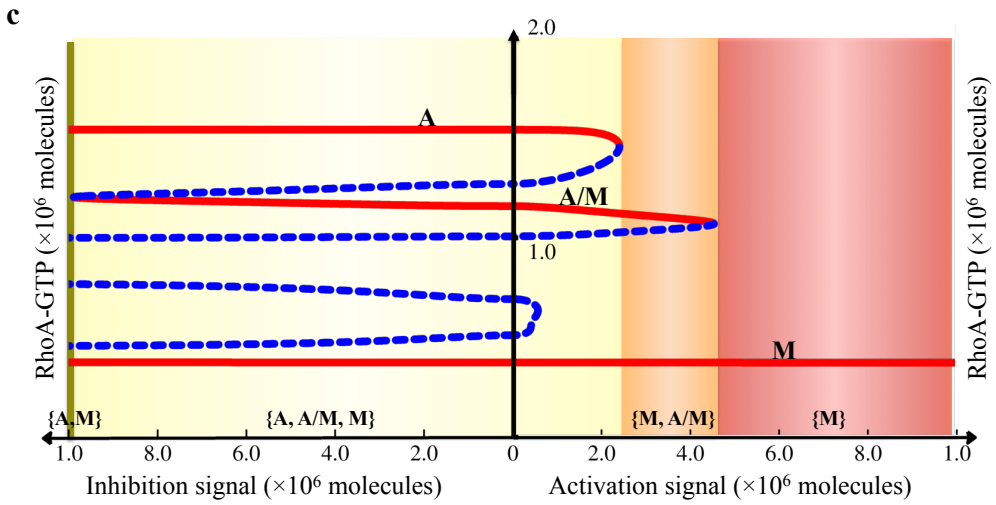


Figure S1.1 - The circuit connecting RhoA and Rac1 to Gab1 and Grb2 signals.

Grb2 can accelerate the GTP loading rate of Rac1 through Ras with the help of Sos, PI3K. When Grb2 binds and activate SOS, Ras is then activated[11,13]. The activated Ras can activate Rac1 either in PI3K dependent pathway[11] or by Tiam1 in PI3K independent way[12]. Gab1 can accelerate the GTP loading rate of either Rac1 or RhoA through Crk/DOCK180[14] and LARG[15] respectively.


a
3.0
M
A/M
1.5
E/M
A
{M}
{A,M}
{A, E/M, M}
{A, A/M, M}
{A,M}
{A}
Rac1-GTP (×10⁶ molecules)
Rac1-GTP (×10⁶ molecules)
0.5 0.4 0.3 0.2 0.1 0 0.1 0.2 0.3 0.4 0.5
Inhibition signal (×10⁶ molecules)
Activation signal (×10⁶ molecules)


b
3.0
M
A/M
1.5
A
{A,M}
{A, A/M, M}
{M, A/M}
{M}
Rac1-GTP (×10⁶ molecules)
Rac1-GTP (×10⁶ molecules)
1.0 0.8 0.6 0.4 0.2 0 0.2 0.4 0.6 0.8 1.0
Inhibition signal (×10⁶ molecules)
Activation signal (×10⁶ molecules)


c
2.0
A
A/M
1.0
M
{A,M}
{A, A/M, M}
{M, A/M}
{M}
RhoA-GTP (×10⁶ molecules)
RhoA-GTP (×10⁶ molecules)
1.0 8.0 6.0 4.0 2.0 0 2.0 4.0 6.0 8.0 1.0
Inhibition signal (×10⁶ molecules)
Activation signal (×10⁶ molecules)

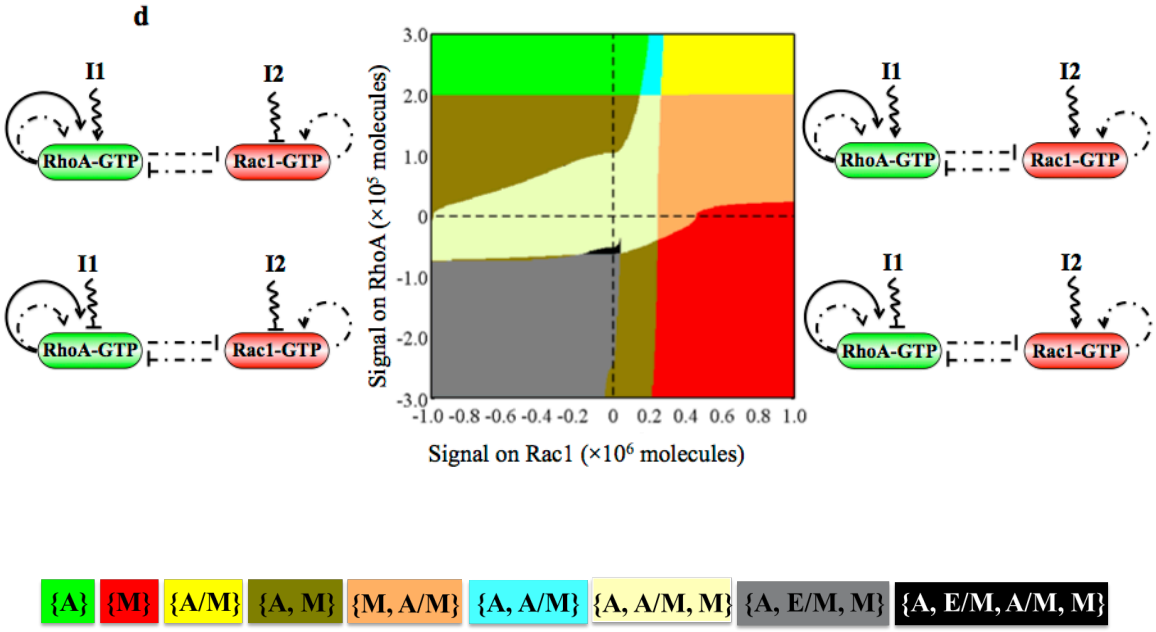


Figure S1.2 - Bifurcations and phase diagram of the circuit when driven by external signals.

(**a)** The complementary bifurcation for Figure 2.4b. Here it is showed in the term of Rac1-GTP protein level. **(b)** Bifurcation of Rac1-GTP protein levels when driven by external signal on Rac1. **(c)** Bifurcation of RhoA-GTP protein levels when driven by external signal on Rac1. The red solid line stands for stable states and the blue dash line stands for unstable states. **(d)** Phase diagram of external signals both on RhoA (I_1) and Rac1 (I_2). The positive values stand for activation signals, while the negative values represent the inhibition signals. The circuit for each quadrant is showed beside it. Different coexistences (Phases) are highlighted by different background colors, as illustrated at bottom.

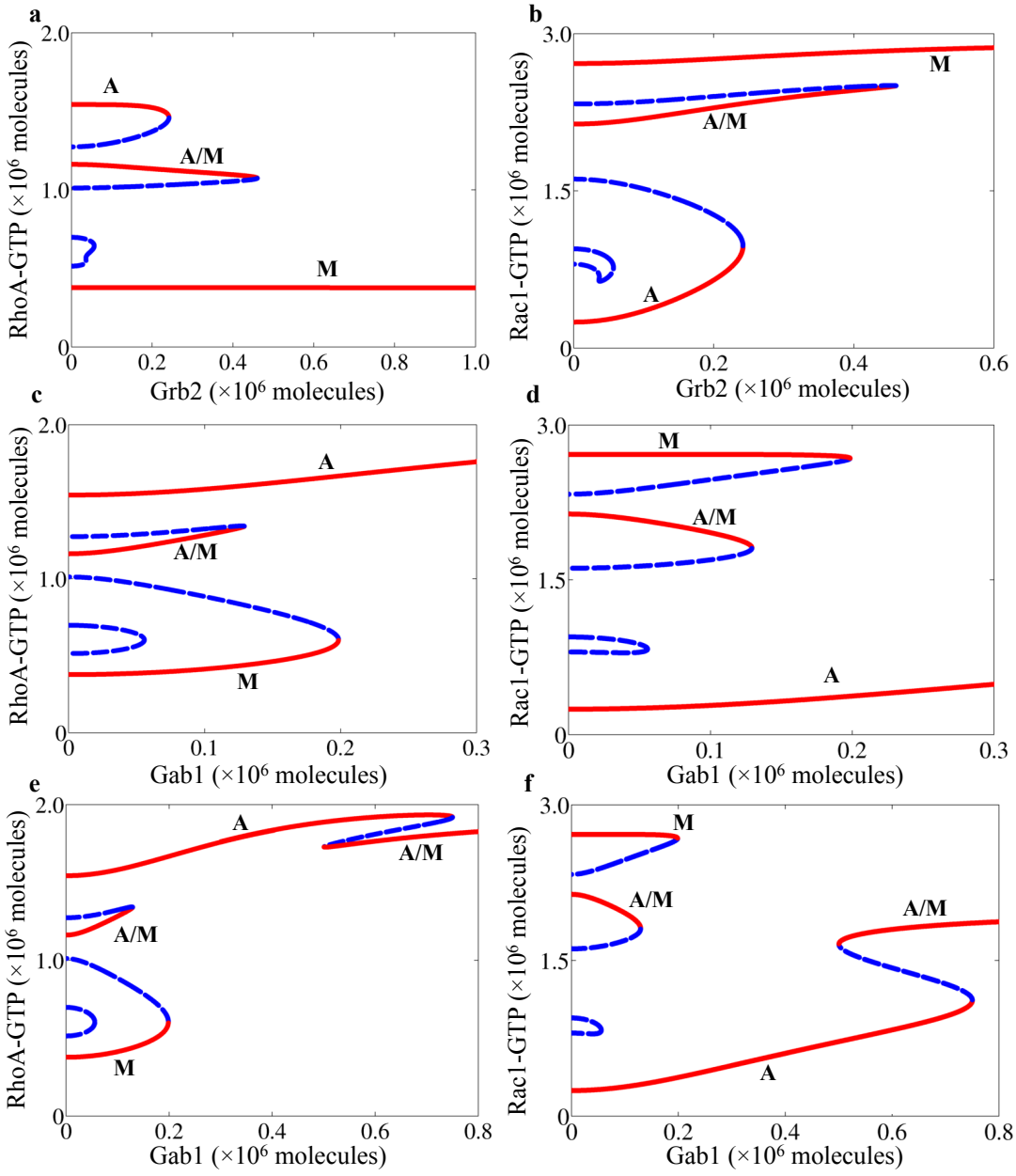


Figure S1.3 - One-parameter bifurcation diagrams for Grb2 and Gab1 regulation.

(a) Bifurcation diagrams as the function of Grb2 and Rac1-GTP. **(b)** Bifurcation diagrams as the function of Grb2 and RhoA-GTP. **(c)** Bifurcation diagrams as the function of Gab1 and Rac1-GTP. **(d)** Bifurcation diagrams as the function of Gab1 and RhoA-GTP. When Gab1 signal increase further, the cells finally are induced to A/M phenotype (**e, f**). Blue dashed lines stand for the unstable states, while red solid lines

stand for the stable states. The phenotypes corresponding to each stable line are labeled in the figures.

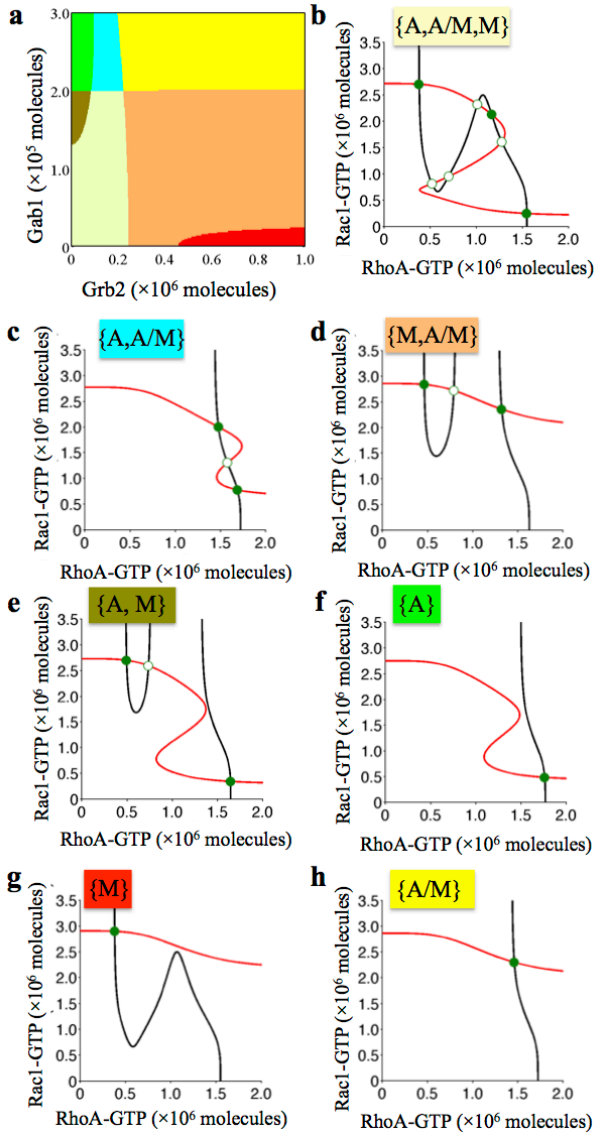


Figure S1.4 - Details about the different phases in Grb2/Gab1 phase diagram for Rac1/RhoA regulatory circuit.

(a) Phase diagram using Grb2 and Gab1 signals as two parameters (also in Figure 2.5). **(b)** to **(h)** show the nullclines for the circuit at different phases - one phase for tri-stability **(b)**, three phases for bi-stability **(c, d, e)** and three phases for mono-stability **(f, g, h)**.

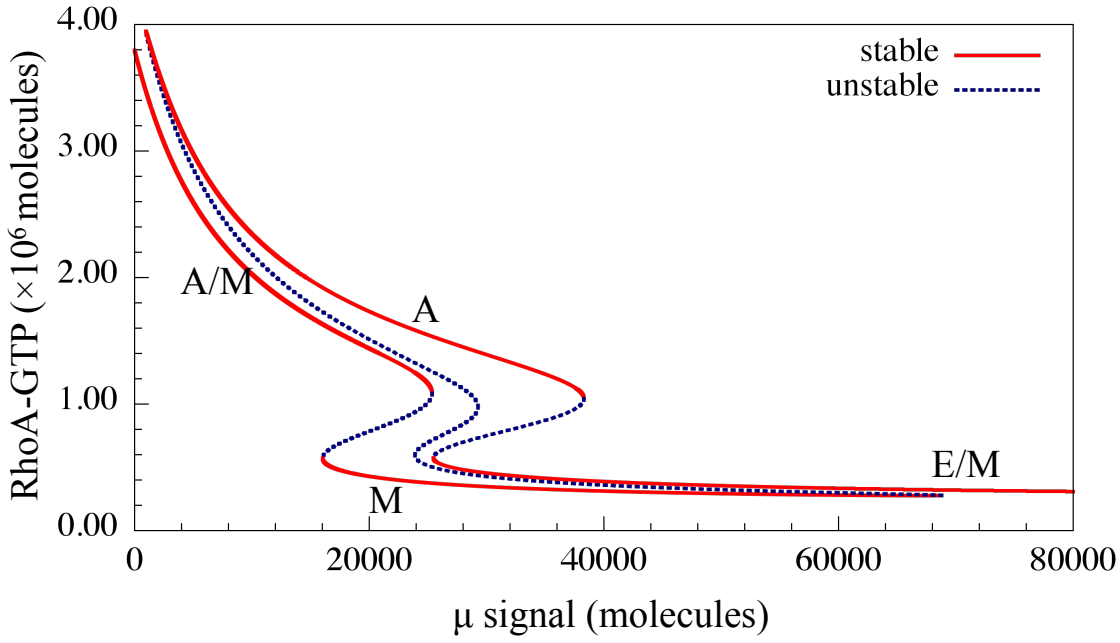


Figure S1.5 - Bifurcation of RhoA-GTP protein levels in response to signal μ.

This diagram is complementary to Figure 2.9, but it is in the term of RhoA-GTP protein level. The red solid lines indicate stable states and the blue dashed lines indicate unstable states.

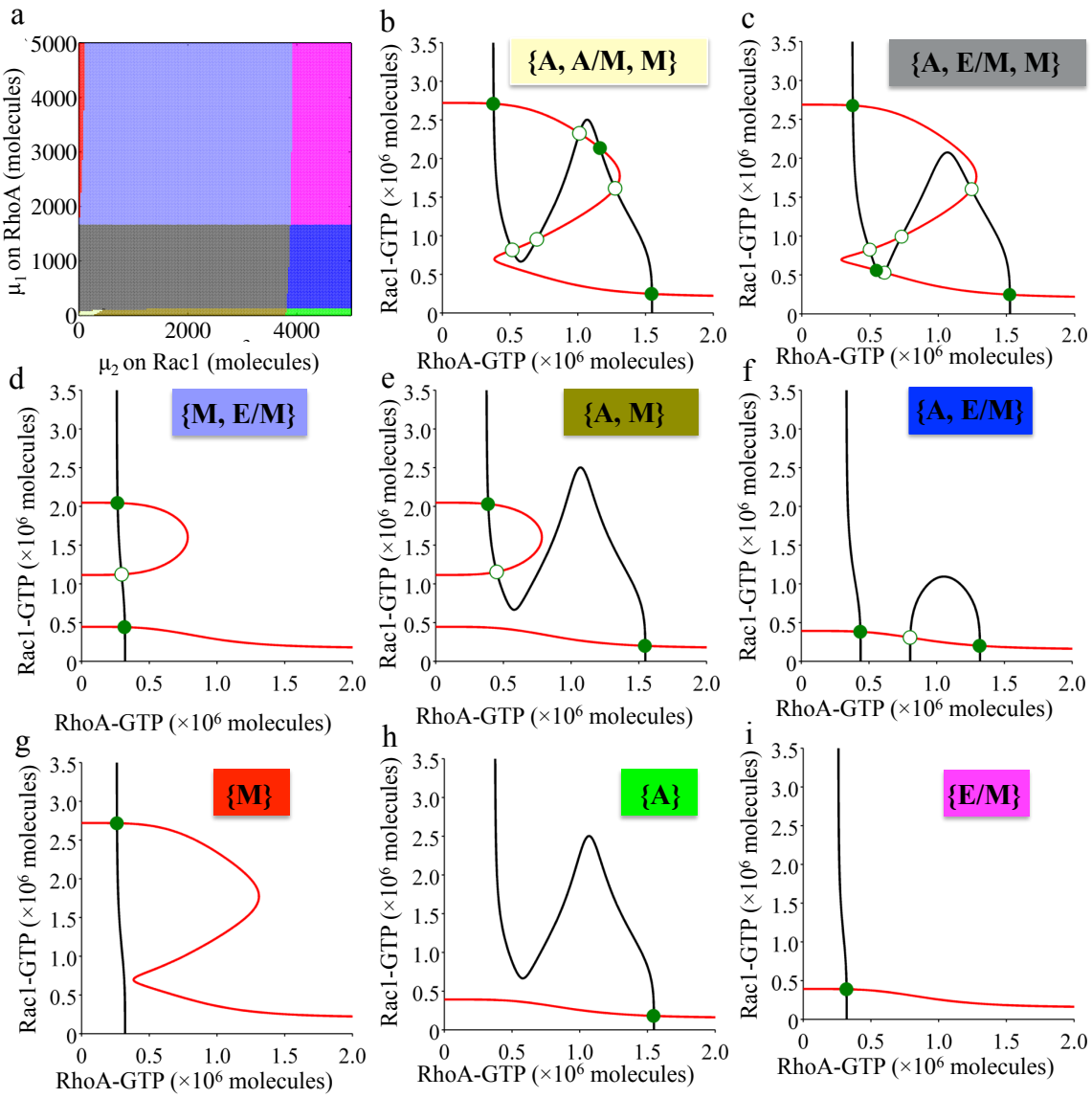


Figure S1.6 - Different phases in μ_1 / μ_2 phase diagram for Rac1/RhoA regulatory circuit.

(a) Phase diagram using μ_1 (y-axis) and μ_2 (x-axis) signals as two parameters (also in Figure 2.10). **(b)** to **(h)** show the nullclines for the circuit at different phases - two phase for tri-stability **(b, c)**, three phases for bi-stability **(d, e, f)** and three phases for mono-stability **(g, h, i)**.

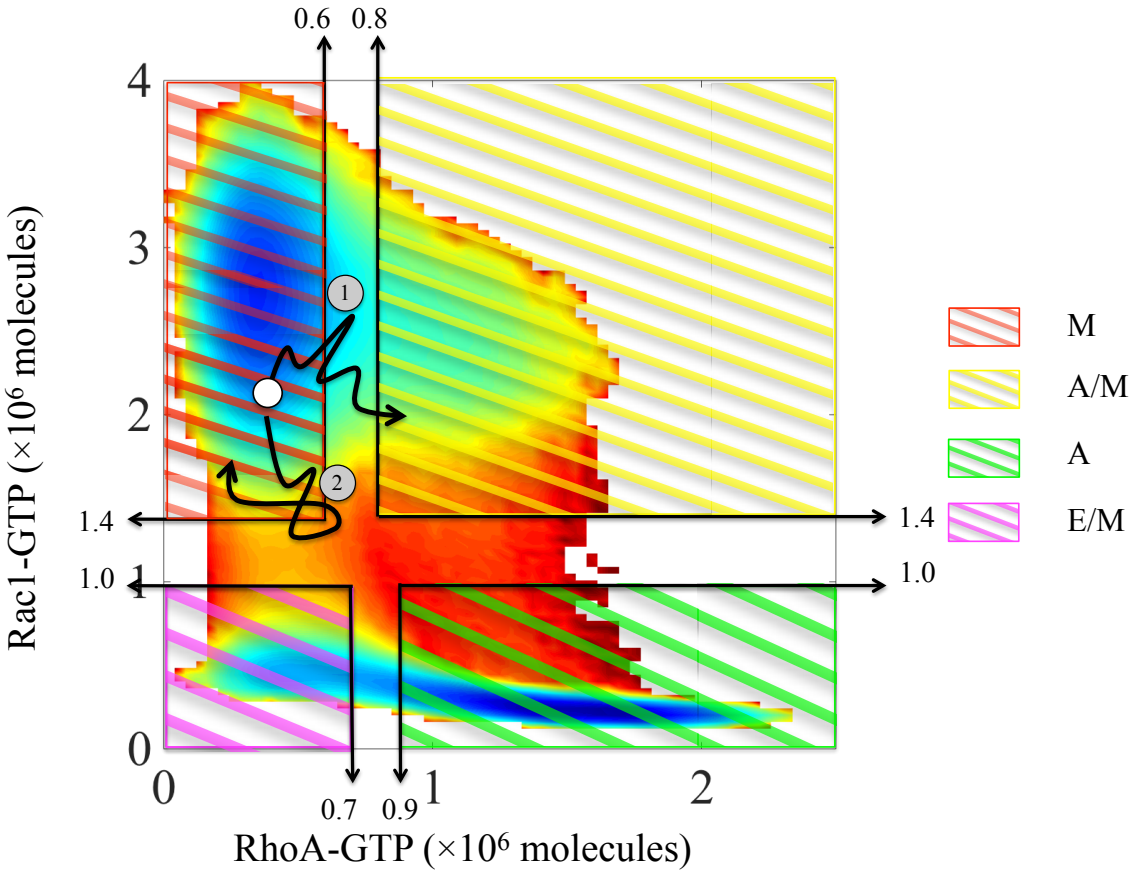


Figure S1.7 - Boundaries for each state. Four different color areas are defined to distinguish the four possible states (E/M, M, A, A/M).

The numbers by the arrow indicate the position of the boundaries. The numbers by the arrow indicate the position of the boundaries. Trajectory 1 and 2 both started from same initial state, but trajectory 1 finally make successful transition from M to A/M while trajectory 2 failed and returned to M state. Based on these boundaries, we could calculate the number of transition per hour for specific transition (Figure. S1.8a) and the probability of specific state (Figure. S1.8b). The probability of specific state is defined as the ratio of accumulating time staying in the state over the total time of the simulation (10^6 hours).

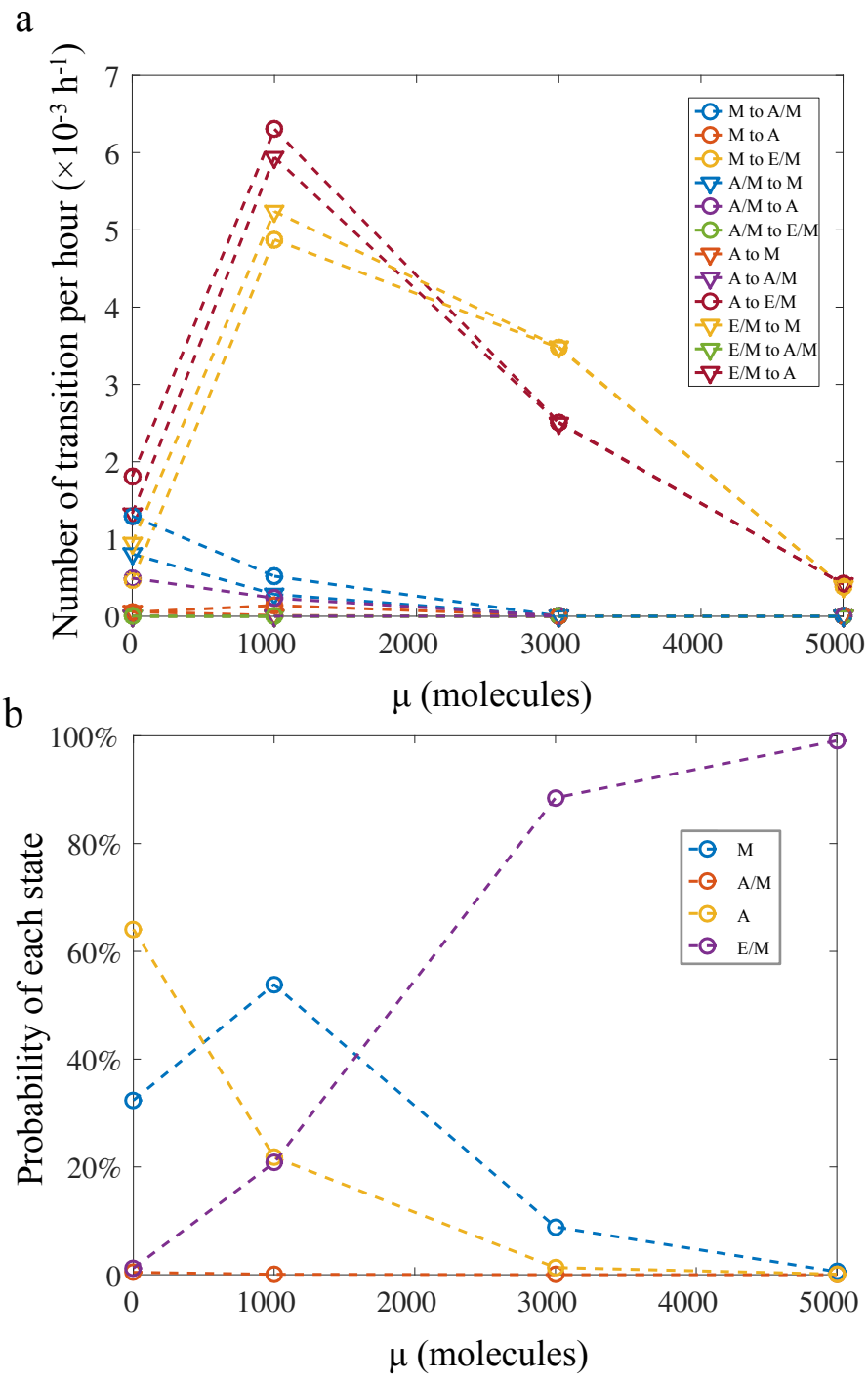


Figure S1.8 - Number of transition per hour between different states and the probability of each state.

(a) At different level of signal μ (0, 1000, 3000, and 5000 molecules), the number of transition per hour for 12 possible transitions are shown by different lines. **(b)** At different level of signal μ, the probability of each state (E/M, M, A/M, A) is shown by different lines.

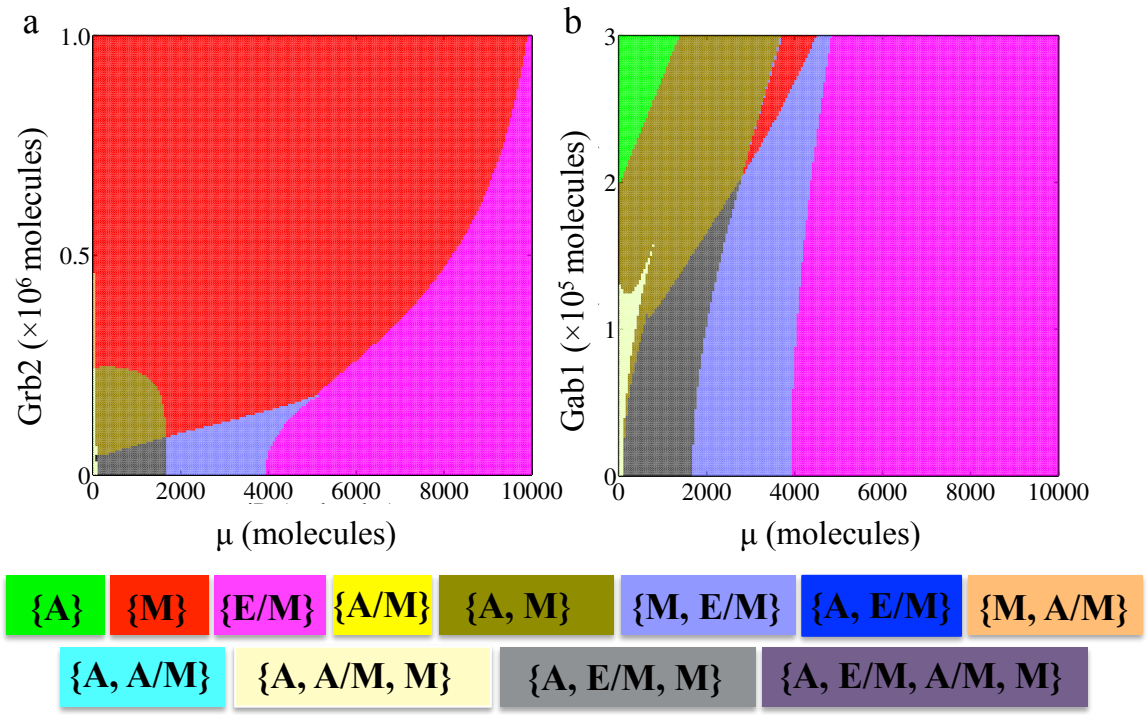


Figure S1.9 - Two-parameter bifurcation diagram for signal μ and signals from c-Met pathway.

(a) Diagram for μ and Grb2, which could activate Rac1 protein. **(b)** Diagram for μ and Gab1, which could activate both Rac1 and RhoA. The colors for different phases are shown at the bottom.

Appendix 2: Development of Random Circuit Perturbation (RACIPE) method to interrogate the topological robustness of gene regulatory circuits

Table S2.1 - Ranges of the parameters for randomization

Parameter	Range[#]
Maximum production rate (G)	1-100
Degradation rate (k)	0.1-1
Fold change (λ)[*]	1-100
Threshold (X_0)	Depend on inward regulations
Cooperativity of the regulation (n)	1, 2, 3, 4, 5, 6

[#]Units: the range of values for each parameter may have different meanings/units for different organisms. For example, the maximum production rate (G) can be 10^2 – 10^4 copies per hour per cell for E. *Coli*, and it can be 10^3 - 10^5 copies per hour per cell for H. *sapiens*. The degradation rate (k) can be 0.1 – 1 hour^{-1} for E. *Coli*, and it can be 0.01 – 0.1 hour^{-1} for H. *sapiens* (data inferred from bionumbers.org[16]). In the current RACIPE method, the exact unit is not crucial to the results, because the gene expression data for all the RACIPE models were normalized before further analysis.

[*]For inhibition, fold change ranges from 0.01 to 1, but a uniform distribution is sampled for the inverse of λ.

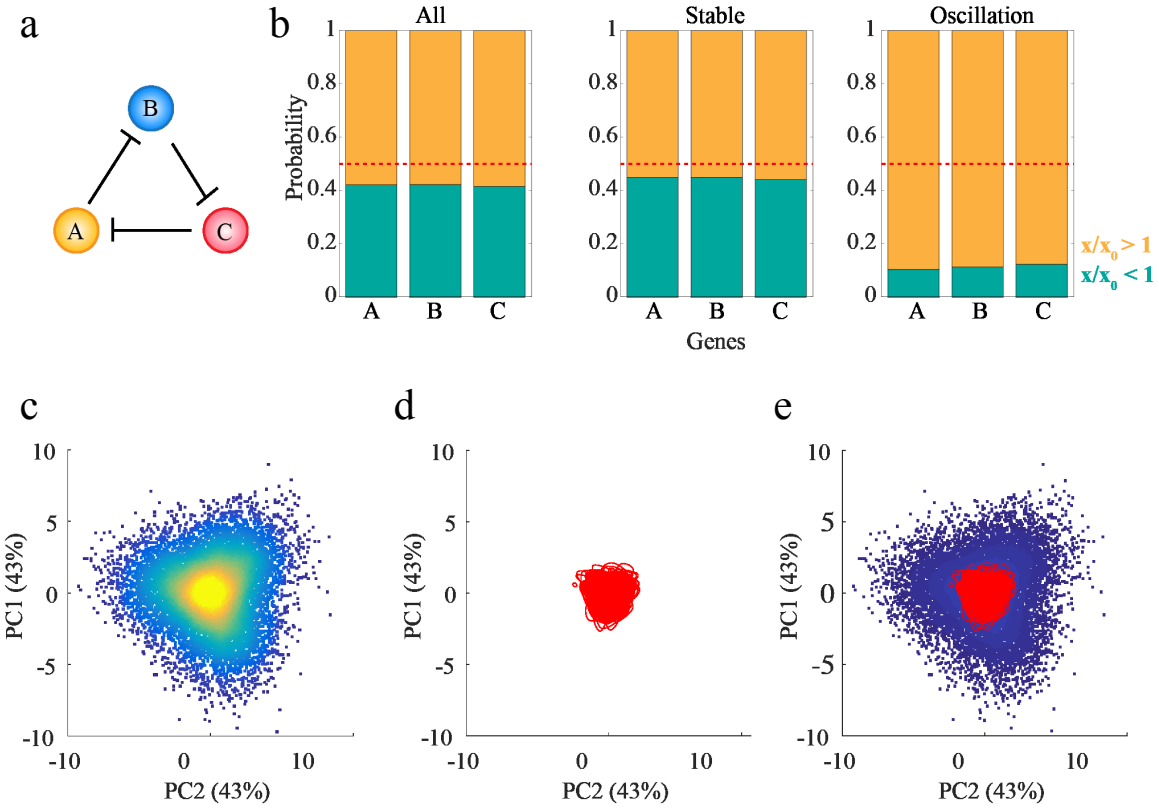


Figure S2.1 -Application of RACIPE to study oscillatory dynamics of a repressilator gene circuit.

(a) Illustration of a repressilator circuit with three genes, where each gene represses the next gene in the circuit. **(b)** Tests of the half-functional rule for all RACIPE models (leftmost panel), the models with stable steady states (middle panel), and the models with stable oscillation (right panel). For the models with stable oscillation, we computed the ratio (x/x_0) of the mean level of each gene X during oscillation and the threshold (x_0) for the outward regulations from gene X. The yellow region shows the probability of $x/x_0 > 1$ for the models, and the green region shows the probability of $x/x_0 < 1$. **(c)** 2D probability density map of the RACIPE-predicted gene expression data of the models with stable steady states projected to the 1st and 2nd principal component axes. **(d)** Projection of the oscillatory trajectories of the models with stable oscillations to the same 1st and 2nd principal component axes in (c). **(e)** The overlapping of the PCA results between (c) and (d).

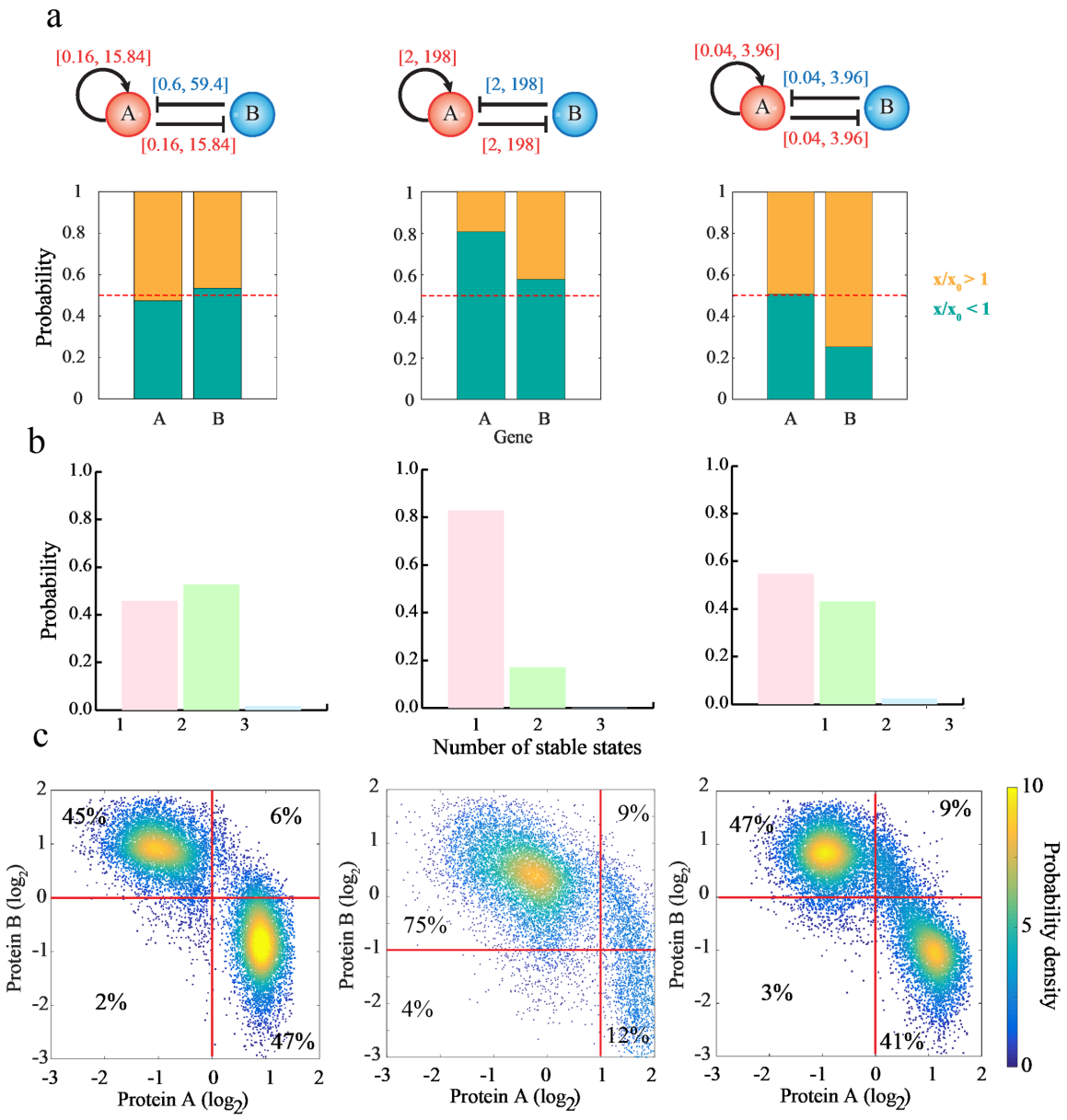


Figure S2.2 - Tests of several random sampling schemes with and without the half-functional rule.

(a) Test of the half-functional rule of a toggle-switch with one-sided self-activation where different ranges were used to randomize the threshold parameters. The leftmost panel shows the circuit and the sampling ranges of the threshold parameters by RACIPE according to the half-functional rule. The middle and the rightmost panels show two examples where same ranges of threshold parameters are used for all regulations. **(b)**

Probability distributions of the number of stable steady states for each circuit. **(c)** Probability density maps of the gene expression data from all the RACIPE models, where the fraction of RACIPE models in each quadrant is shown.

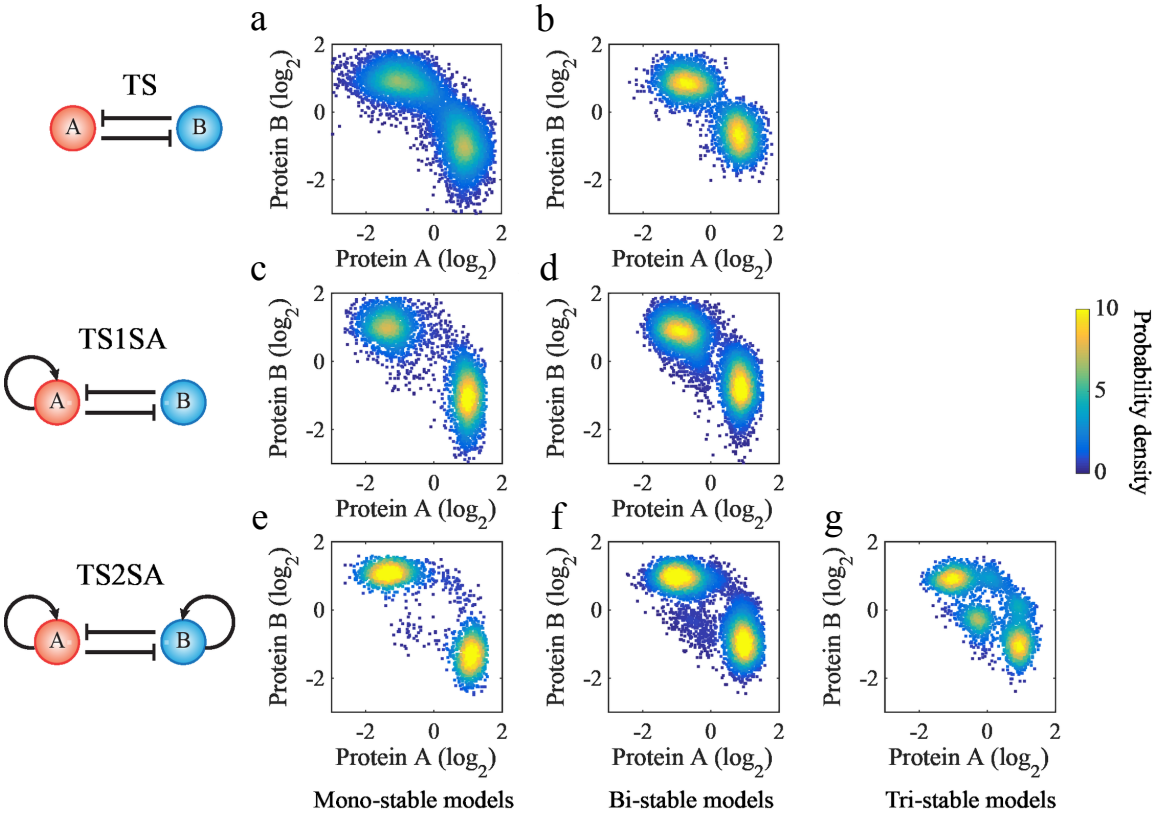


Figure S2.3 - Probability density maps of the gene expression data from all the RACIPE models with a fixed number of stable states for TS (a, b), TS1SA (c, d) and TS2SA (e, f and g) motifs.

a, c and e are the maps for mono-stable models, b, d, and f are the maps for bi-stable models while g is for the tri-stable models. For models with different number of gene states, the gene state clusters remain the same, but the percentage of models in each gene state varies.

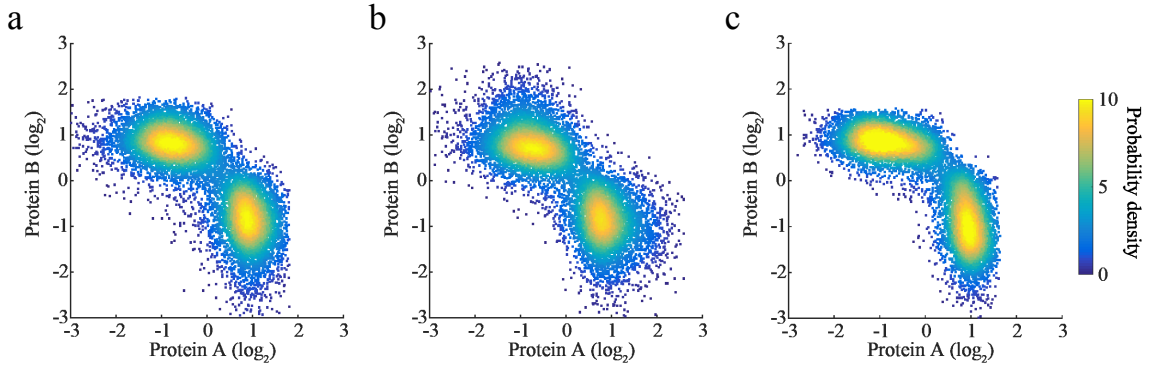


Figure S2.4 - The effects of the range of parametric perturbations on the robust gene states of a toggle-switch circuit motif.

Each panel shows the probability density map of the gene expression data from all the RACIPE models for a version of RACIPE with different range of parametric perturbations. **(a)** The range of the production rates was randomized from 1-1000. **(b)** The range of the degradation rates was randomized from 0.1-10. **(c)** The range of the fold changes was randomized from 1-1000. In each case, the range of variations is 10 times as large as the original method. The sampling distribution for parameter randomization for these cases is still a uniform distribution.

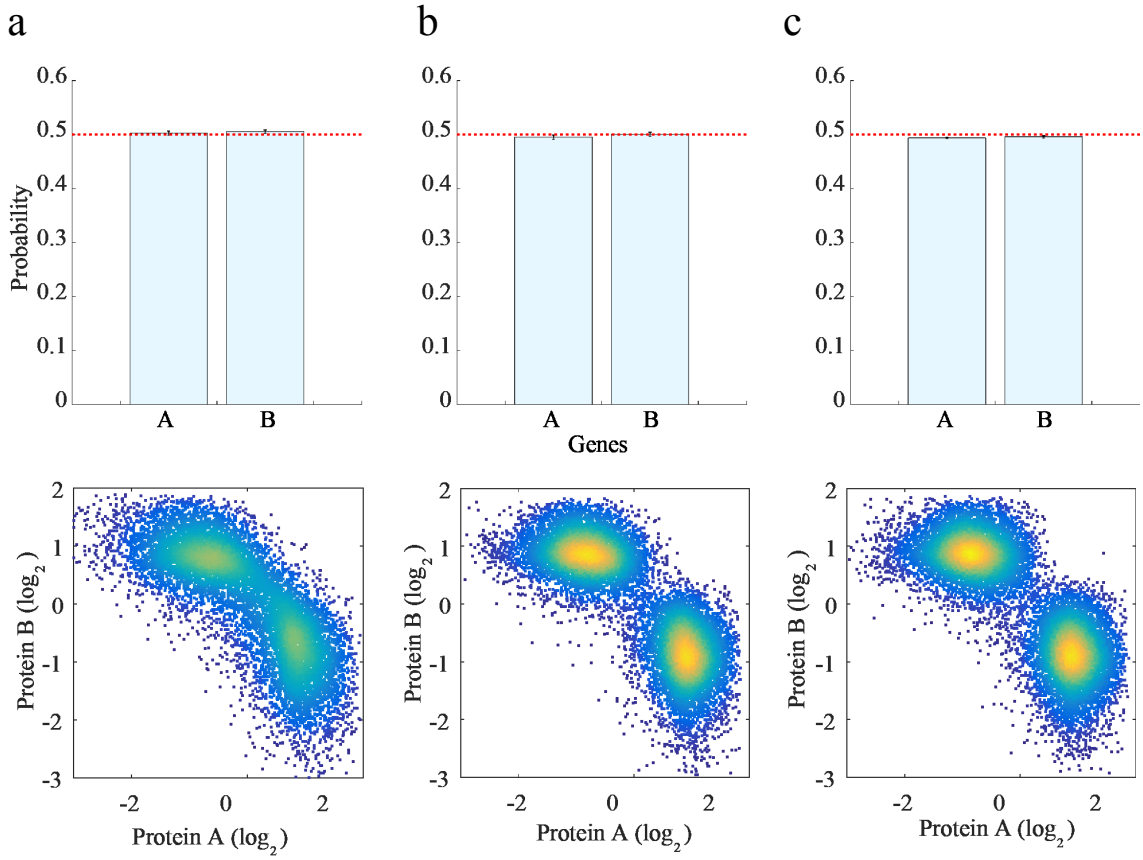


Figure S2.5 - The effects of Hill coefficients on the robustness of the sampling scheme and gene states of a toggle-switch circuit motif.

Test of the half-functional rule (top-panels) and 2D probability density map (bottom-panels) of RACIPE-generated gene expression data are shown for cases where Hill coefficients are randomized with different ranges -- a: 1-3; b: 4-6; c: 7-9.

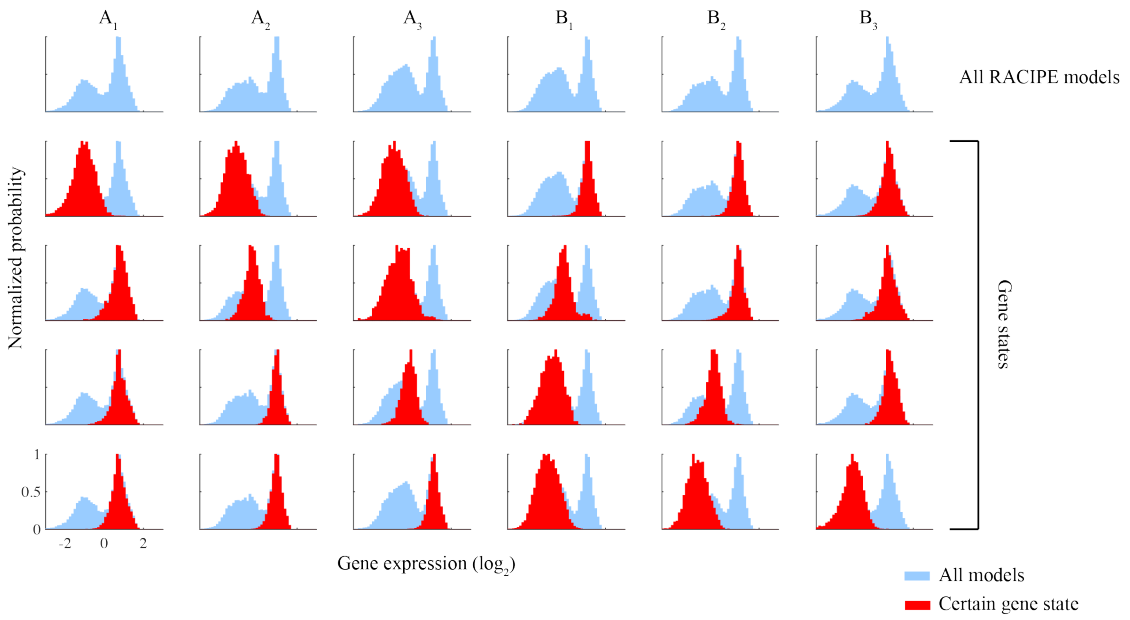


Figure S2.6 - Gene expression distributions of each gene state for the CTS motif with three coupled toggle switches.

The gene expression distribution of each gene for all of the RACIPE models is shown in blue, while that for each gene state is shown in red (50 bins for the histogram of each distribution). Below, each row shows the distribution of each gene for every gene state, listed in the same order as Figure. 3.5b. For clarity, each distribution is normalized by its maximum probability. Each column represents a gene and each row represents a gene state. For each state, the expression of a gene could be assigned as a high, intermediate or low level according to the relative location of its distribution (red) with respect to the distribution (blue) for all the RACIPE models.

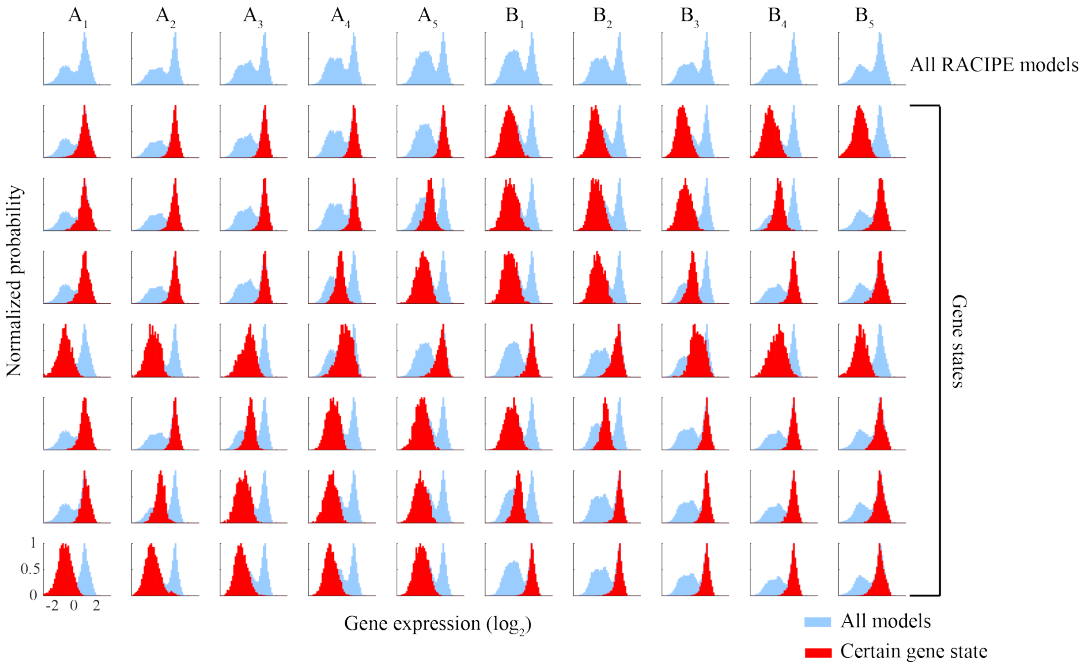


Figure S2.7 - Gene expression distributions of each gene state for the CTS motif with five coupled toggle switches.

The gene expression distribution of each gene for all of the RACIPE models is shown in blue, while that for each gene state is shown in red (50 bins for the histogram of each distribution). Below, each row shows the distribution of each gene for every gene state, listed in the same order as Figure 3.5b. For clarity, each distribution is normalized by its maximum probability. Each column represents a gene and each row represents a gene state. For each state, the expression of a gene could be assigned as a high, intermediate or low level according to the relative location of its distribution (red) with respect to the distribution (blue) for all the RACIPE models.

Appendix 3: Decoding Stem Cell Regulatory Circuit by Random Circuit Perturbation Method

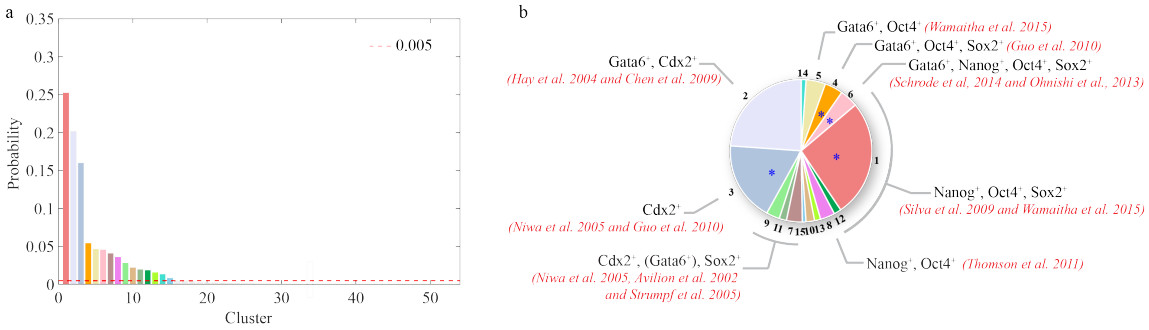


Figure S3.1 - Identification of major gene states for the stem cell gene regulatory circuit.

(a) A total of 54 gene clusters were identified by the hierarchical average linkage clustering analysis on the RACIPE gene expression data for the stem cell circuit in Figure 4.1b. With a minimum probability cutoff of 0.005, we identified 15 clusters, referred to as major gene states. The coloring scheme for these 15 clusters is consistent with that used in Figure 4.1b, and the other clusters are shown in grey. **(b)** Characteristic gene expressions and experimental evidences for each of these gene states, ranked by the likelihood in the RACIPE models as in (a). The four gene states that match the single cell experimental data (Figure 4.1b, right) are highlighted by blue asterisks and their likelihoods.

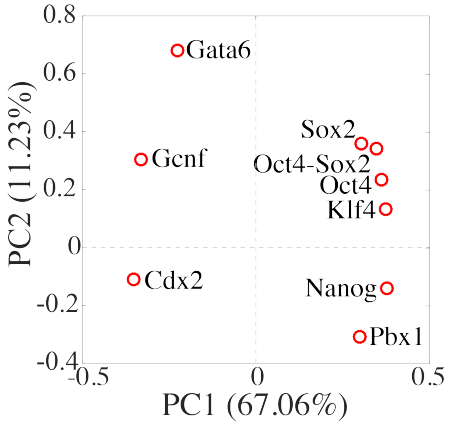


Figure S3.2 - Contribution of each gene to the 1st and 2nd principal component axes.

The principle component axes were obtained by the PCA of the gene expression data from all of the RACIPE models. The position of each gene in the diagram illustrates its contribution to the two axes. *Oct4/Sox2/Nanog* have position contributions to the first axis, while *Cdx2* and *Gata6* have negative contributions to the first axis.

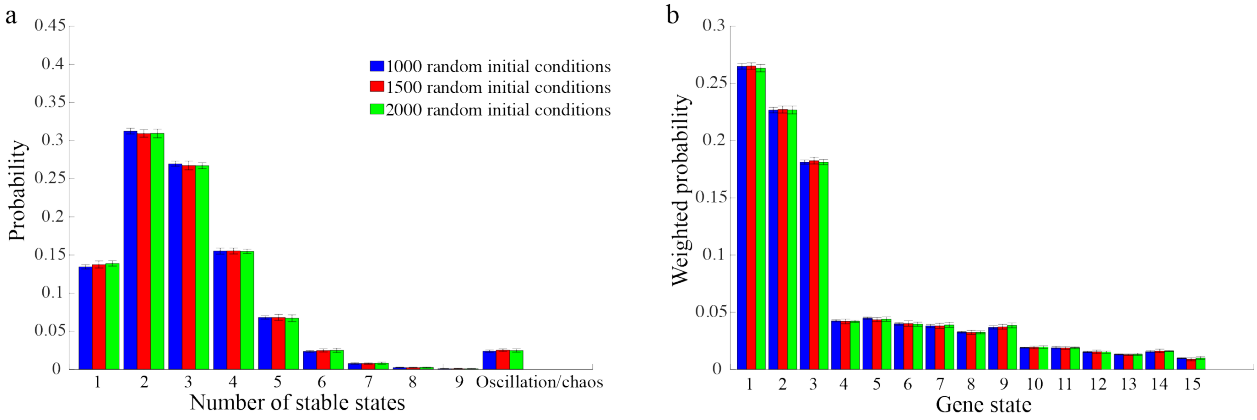


Figure S3.3 - The effects of the number of random initial conditions to the convergence of the RACIPE results.

For each RACIPE model, all the possible stable steady states are obtained by running ODE simulations starting from different random initial conditions. Here, we checked whether the statistical results converged by varying different numbers of initial conditions (blue: 1000 times, red: 1500 times, and green: 2000 times). Each case was repeated 10 times to estimate the mean and the standard deviation of the histogram. **(a)** Probability distribution of the number of stable steady states for each RACIPE model. **(b)** Probability distribution of the gene states.

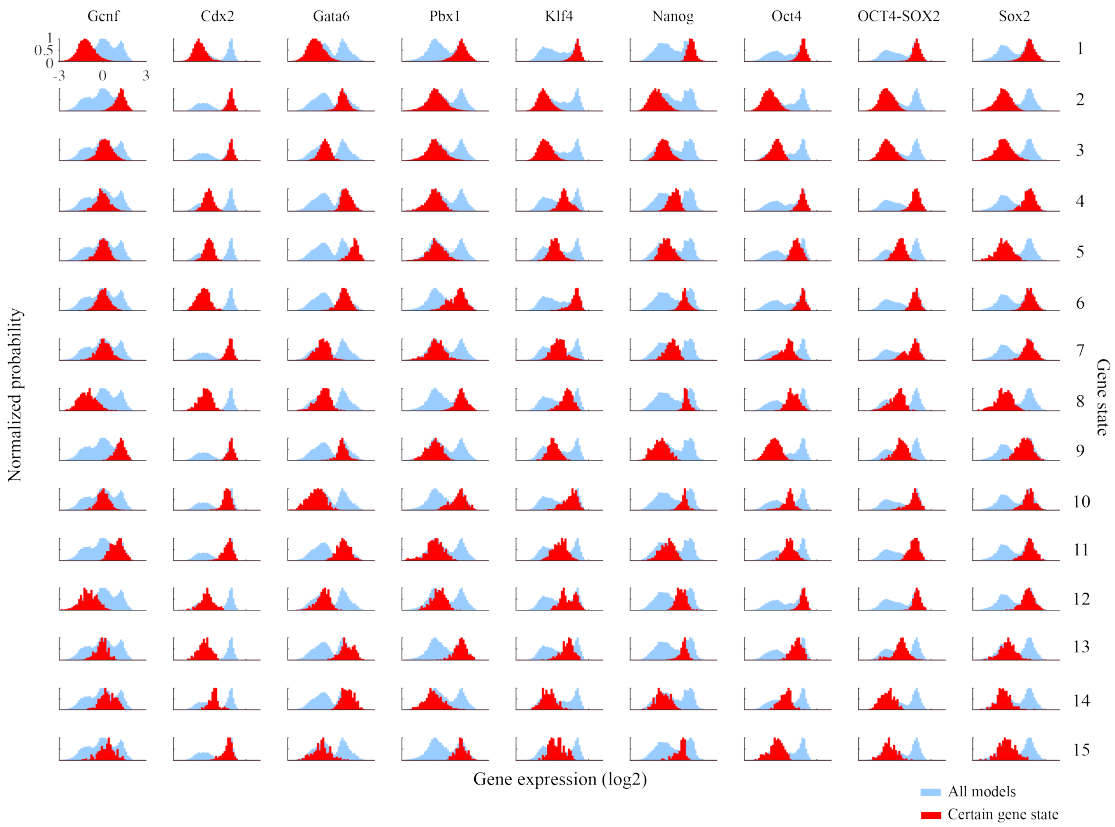


Figure S3.4 - Gene expression distributions of each gene state.

The gene expression distribution of each gene for all of the RACIPE models is shown in blue, while that for each gene state is shown in red (50 bins for the histogram of each distribution). For clarity, each distribution is normalized by its maximum probability. Each column represents a gene and each row represents a gene state. The results illustrate that RACIPE, unlike the traditional Boolean network model[17], allows the identification of the intermediate expression for some gene states.

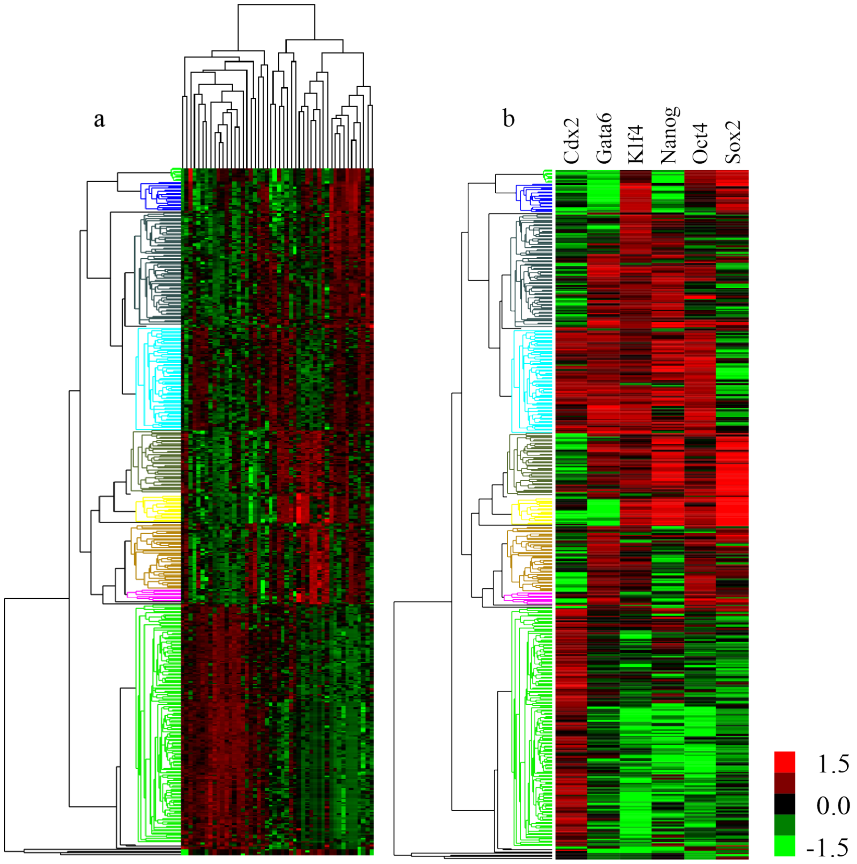


Figure S3.5 - Hierarchical clustering of mouse embryo single cell gene expression data.

(a) Clustering result of a total of 48 genes[18]. **(b)** The extracted gene expression data for 6 genes (Cdx2, Gata6, Klf4, Nanog, Oct4 and Sox2). Each column represents a gene while each row represents a cell.

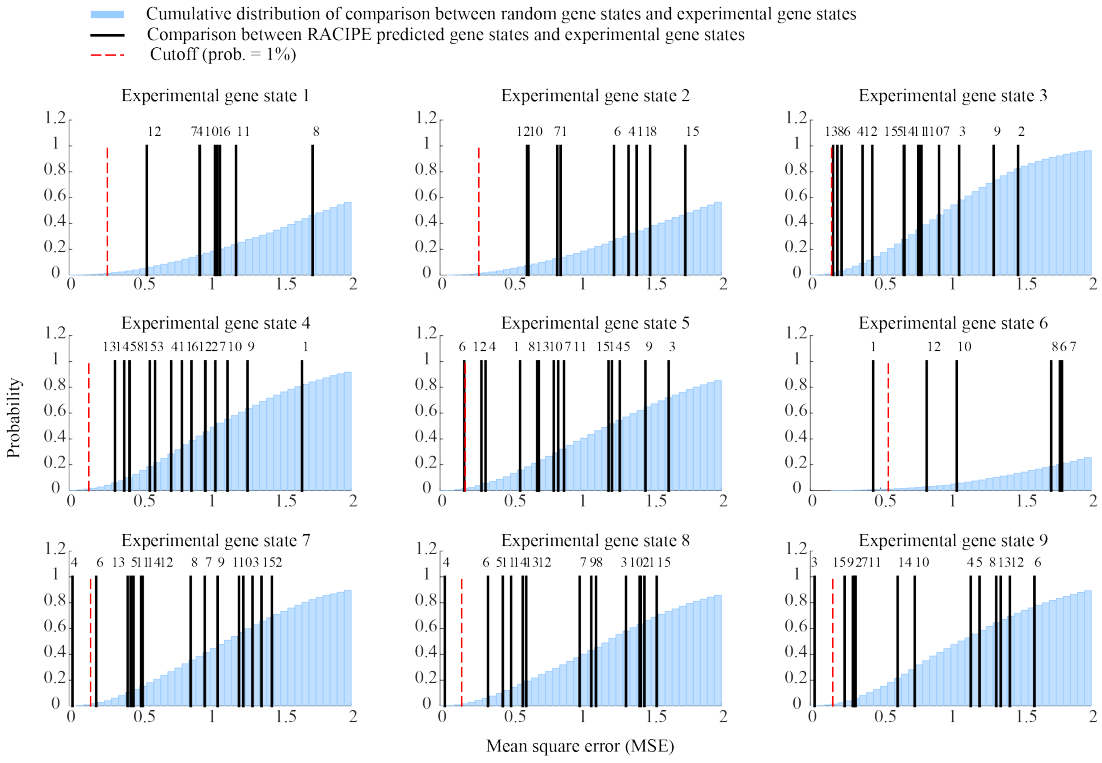


Figure S3.6 - Comparison between the experimental gene states and the RACIPE gene states.

The comparison analysis was done by the cluster-based analysis (See Methods). First, we calculated the average gene expression vector of each experimental cluster (experimental gene state) and each RACIPE cluster (RACIPE gene state). Then, for each experimental gene state, we calculated the mean square errors (MSEs, see Methods) between the average gene expression vector of the experimental gene state and that of every RACIPE gene state (shown as solid lines, labeled on the top). The blue histogram is the cumulative probability distribution of the MSEs between the average gene expression vector of the experimental gene state and 10,000 random gene expression vectors. We considered that an RACIPE gene state matches the experimental gene state if the corresponding MSE between them is below 1% of the cumulative probability (the cutoff, shown as red dashed lines). According to the plots, no match was found for the 1st, 2nd, 3rd and 4th experimental gene states. The 5th experimental gene state matches the 6th RACIPE gene state; the 6th experimental gene state matches the 1st RACIPE gene state; both the 7th and the 8th experimental gene states match the same 4th RACIPE gene state; the 9th experimental gene state matches the 3st RACIPE gene state.

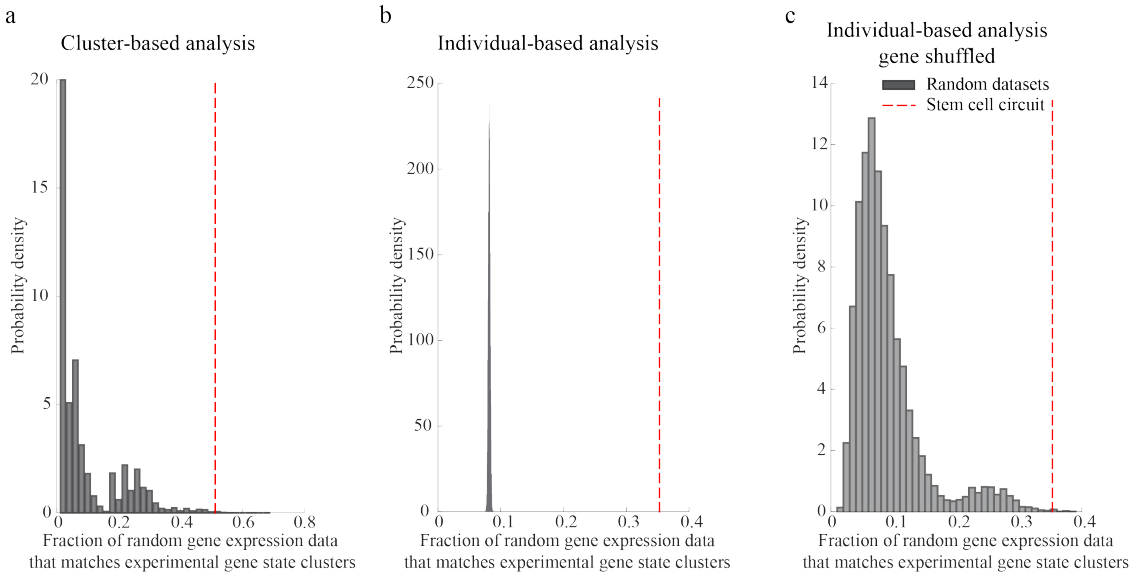


Figure S3.7 - Comparison between experimental gene expression data and random gene expression data.

Random resampling was performed to generate random datasets to calculate the statistical significance (i.e. p-value) of the comparison between experimental data and the RACIPE generated data. Here three different tests were performed. **(a)** Histogram of the fraction of a random dataset that matches the experimental gene states by the cluster-based analysis (see Methods). In this test, we generated 10,000 random datasets, each of which has 15 resampled gene expression vectors representing the average gene expression vectors of a null model (see Methods for the procedure of the resampling). For each random vector, we also randomly assigned the probability of a different RACIPE cluster to the vector. **(b)** Histogram of the fraction of a random dataset that matches the experimental gene states by the individual-based analysis. Here we generated 10,000 random gene expression vectors (same size as the RACIPE dataset) resampled from the distributions of gene expression of each gene in the RACIPE dataset. **(c)** Similar to the individual-based analysis used in (b), but the random datasets were generated in a different way. Here, a random dataset was generated by shuffling the gene identities for each gene state cluster in the RACIPE dataset. For all the tests, we found the RACIPE dataset (red dotted line) outperforms almost all the random datasets (>99%, grey histogram) in matching the experimental dataset.

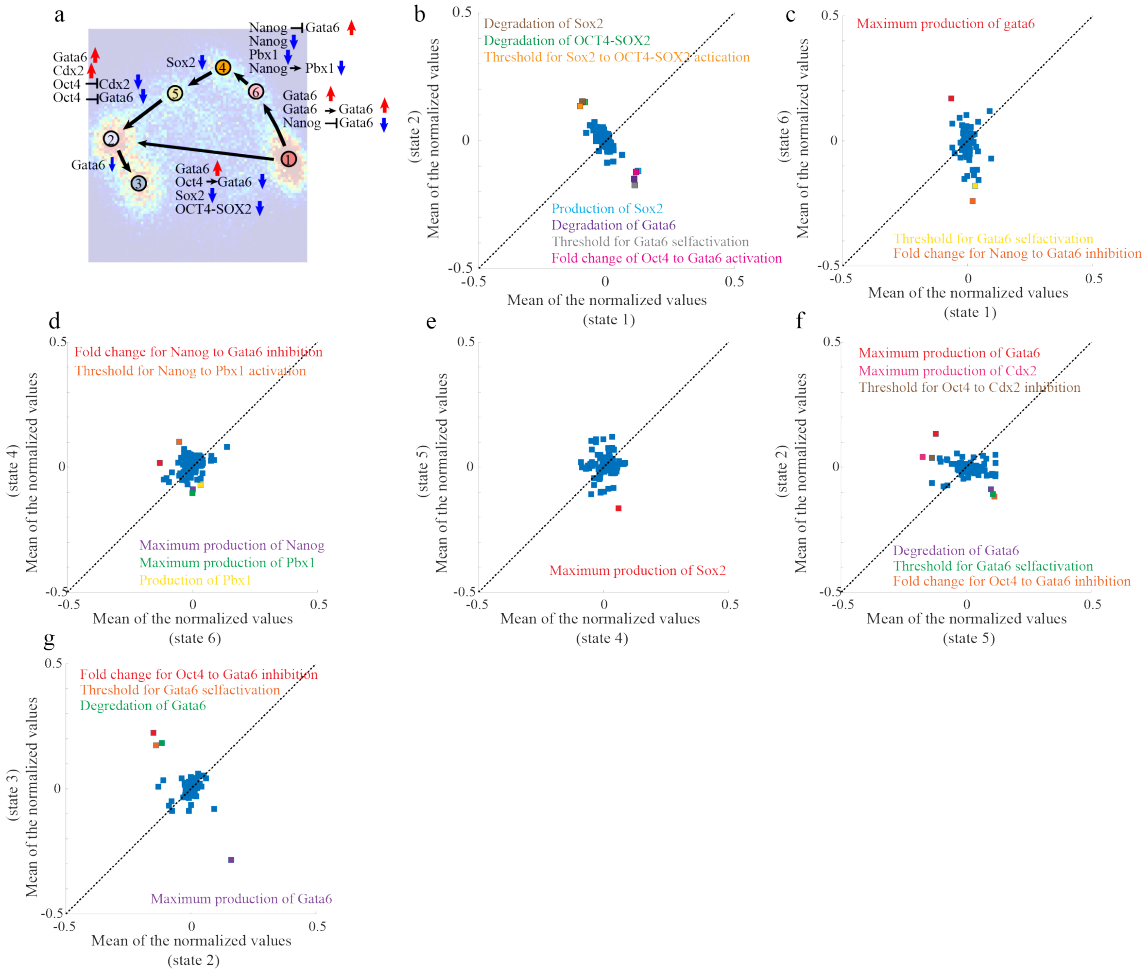


Figure S3.8 - Key parameters that are involved in the transitions among certain gene states.

(a) A summary of the results on top of the probability density map of the RACIPE predicted gene expression data. For each transition shown in panel (a), panels (b) to (g) show the mean of the normalized values for each parameter for the two corresponding gene states (x-axis for the first gene state, and y-axis for the second gene state). **(b)**: states 1 and 2; **(c)**: states 1 and 6; **(d)**: states 6 and 4; **(e)**: states 4 and 5; **(f)**: states 5 and 2; **(g)**: states 2 and 3.

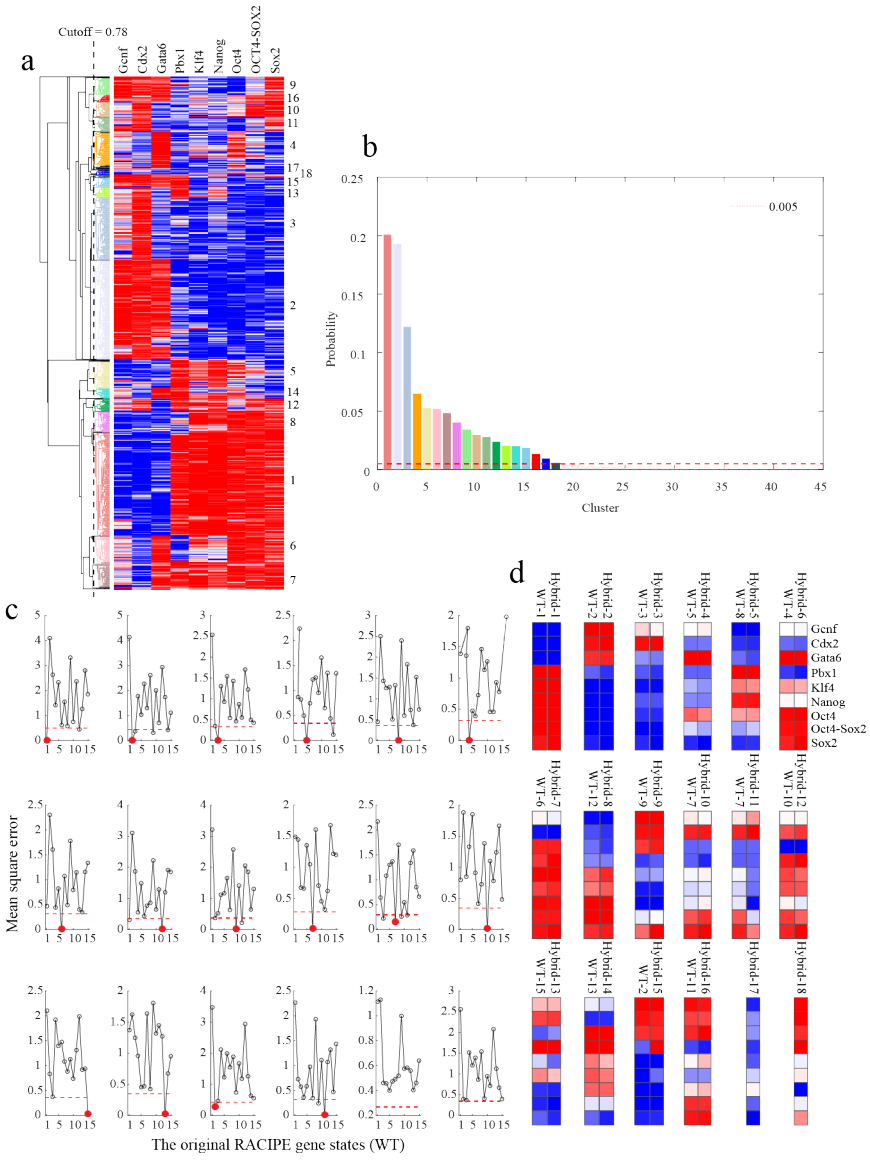


Figure S3.9 - Comparison between the identified gene states from the original RACIPE gene expression data and those from the hybrid RACIPE gene expression data.

(a) Hierarchical average linkage clustering analysis on the hybrid RACIPE gene expression data (Methods). At a cutoff distance of 0.78, 44 clusters were identified. Among them, 18 clusters correspond to 18 major gene states with probability higher than 0.005. (**b**) The probability distribution of all of the gene states highlighted by different colors. **(c)** Comparison between the gene states identified from the hybrid datasets and those identified from the original RACIPE dataset. For each diagram, we computed the

MSEs between the gene expression vectors of a gene state from the hybrid dataset and each of the gene states from the original RACIPE dataset. A cutoff MSE value was selected so that the gene expression vector of the hybrid gene state matches no more than 1% of 10,000 random gene expression vectors. The pairs with the smallest MSE distance below the cutoff (the red dashed line) are colored in red. (**d**) Direct comparison between the matched gene states from the hybrid and the original RACIPE datasets. The heat map shows the average gene expression levels for the corresponding gene states. The results illustrate that the hybrid dataset has similar gene state clusters as the original RACIPE dataset.

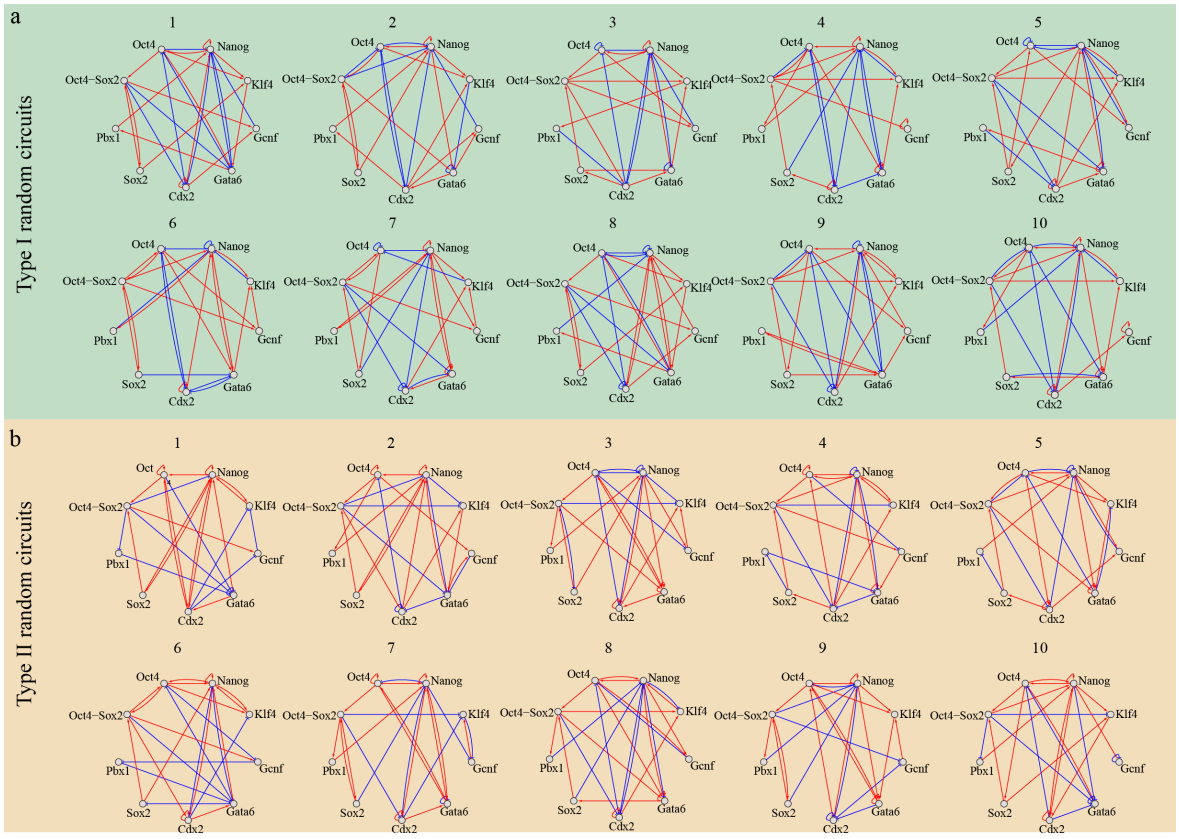


Figure S3.10 - Schematics of random circuits.

(a) Type I random circuits. **(b)** Type II random circuits. Red arrow lines stand for activations and blue bar lines stand for inhibitions. The index for each circuit is shown at the top.

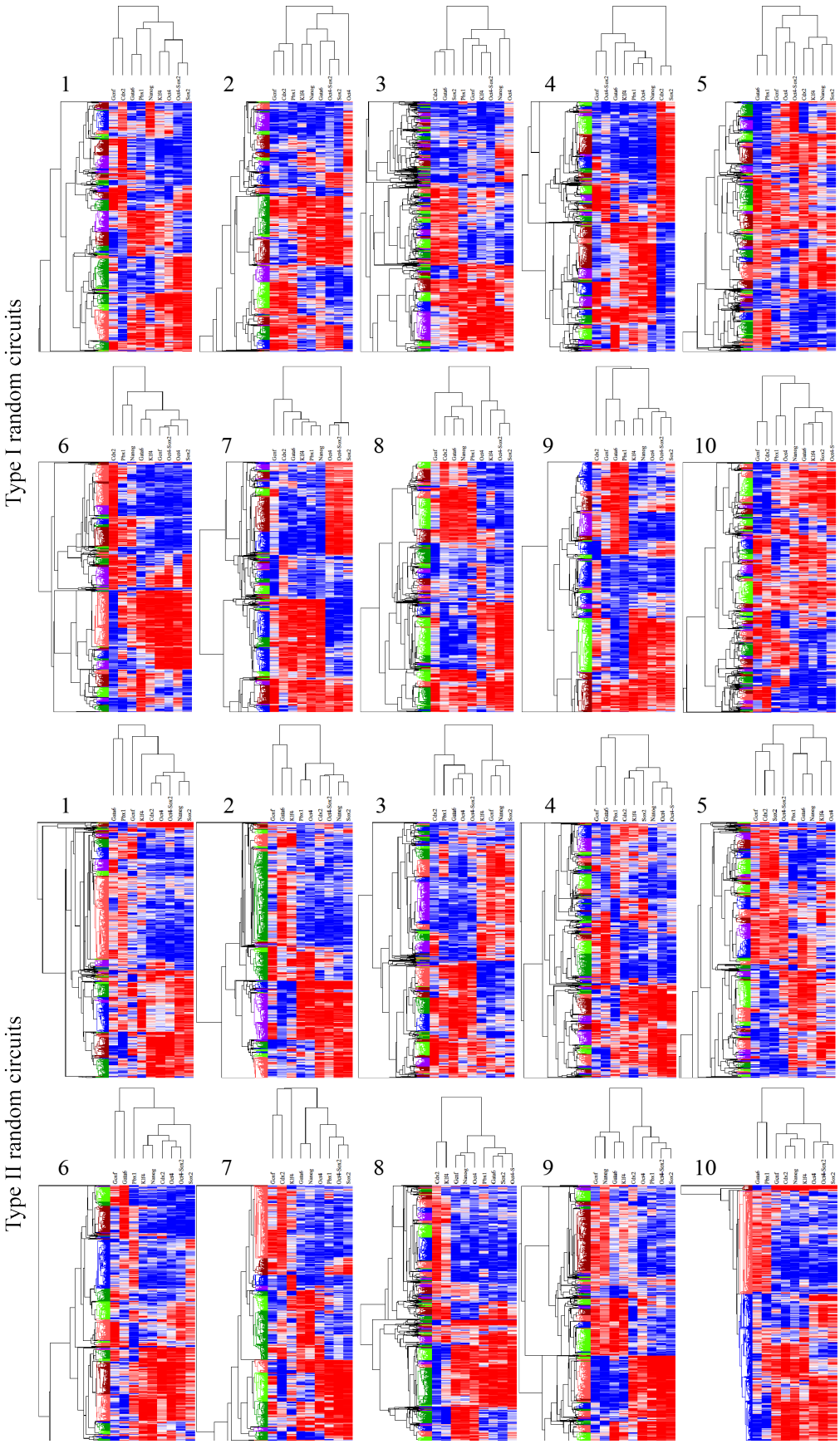

Type I random circuits
Type II random circuits
1
2
3
4
5
6
7
8
9
10

Figure S3.11 - Hierarchical clustering of the RACIPE gene expression data of random circuits.

Each column represents a gene and each row represents a steady state. The cutoff value for the clustering tree is set to be the same 0.8 as that of the stem cell circuit. The identified major gene state clusters are highlighted by colors **for** Type I random circuits **and** Type II random circuits.

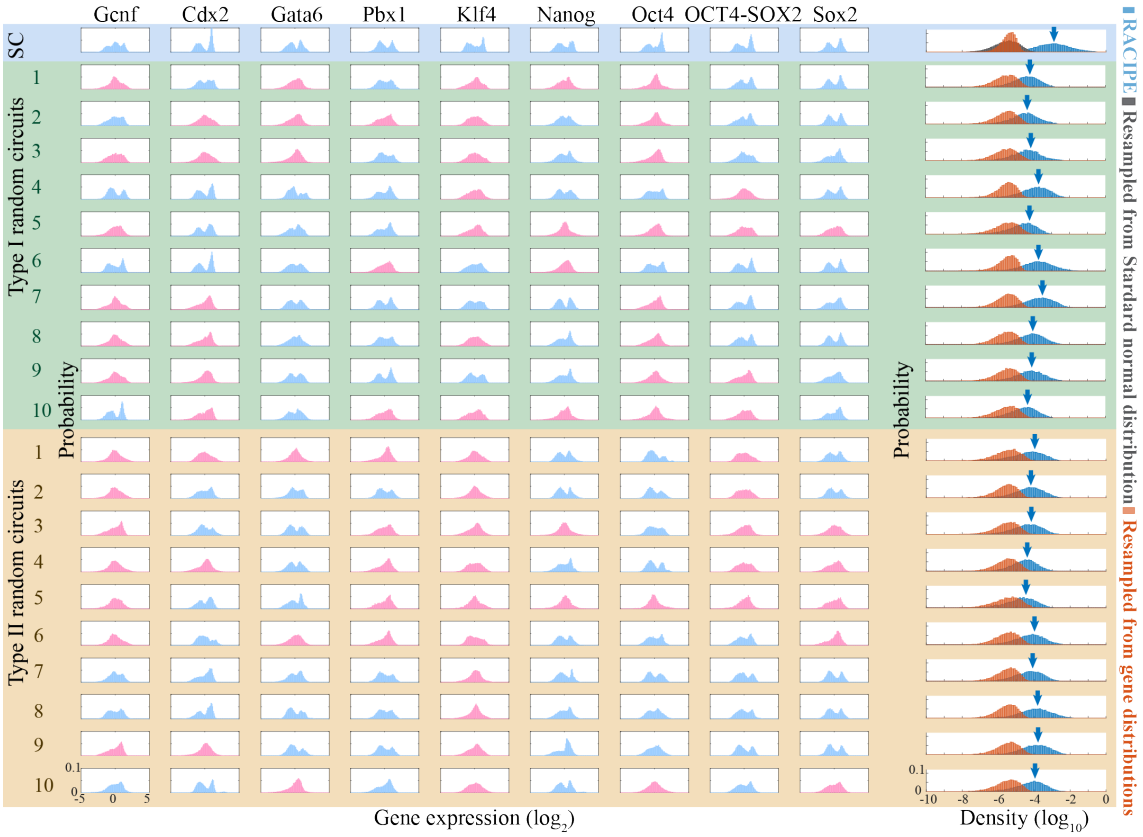


Figure S3.12 - Comparison between the stem cell circuit and a random circuit.

The figure is similar to Figure 4.2, but it includes the results for all of the random circuits. The columns present genes and rows correspond to different circuits. For the results of random circuits, the indices are consistent with those for the random circuits (Figure S3.10).

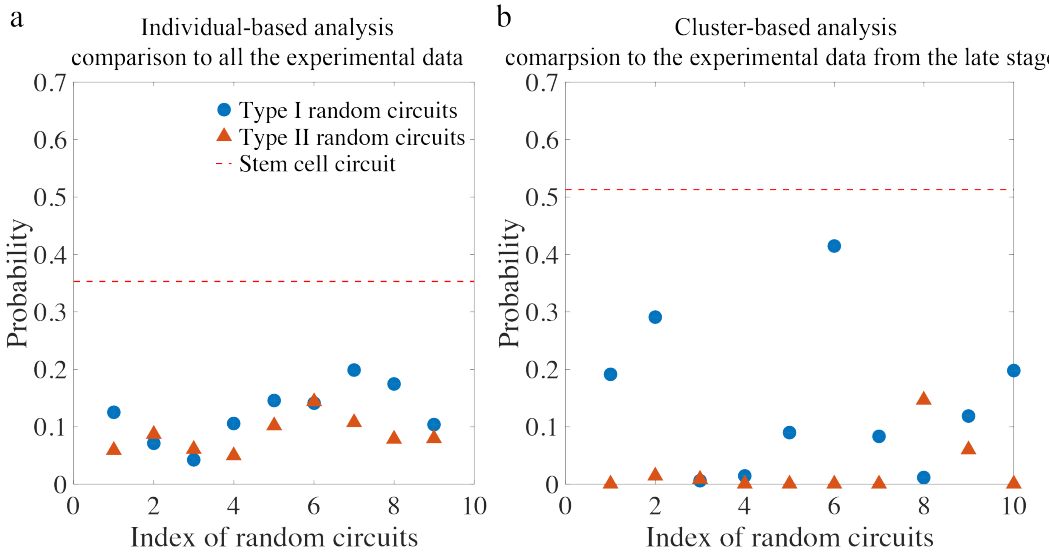


Figure S3.13 - Comparison between the RACIPE gene expression data of a random circuit and experimental gene expression data.

(a) Comparison by the individual-based analysis: RACIPE was applied to each random circuit to generate a collection of gene expression data, which were subject to direct one-on-one comparison with each of the experimental gene expression data (by MSEs, see Methods). The plot shows the probability of the RACIPE gene expression data that matches the experimental data. Blue circuits denote the type I random circuits, and red triangle denote the type II random circuits. The index in the x-axis is consistent with the order in Figure S3.10. **(b)** Comparison by the cluster-based analysis. In addition to the above mentioned procedure, the average gene expression vectors were computed for each gene state cluster from both the RACIPE data and the experimental data. All the RACIPE data from a cluster was regarded to match the experimental data if the average gene expression vector match that of any experimental cluster that belongs to the late stage ($\geqslant$ 32 cells) of embryo development.

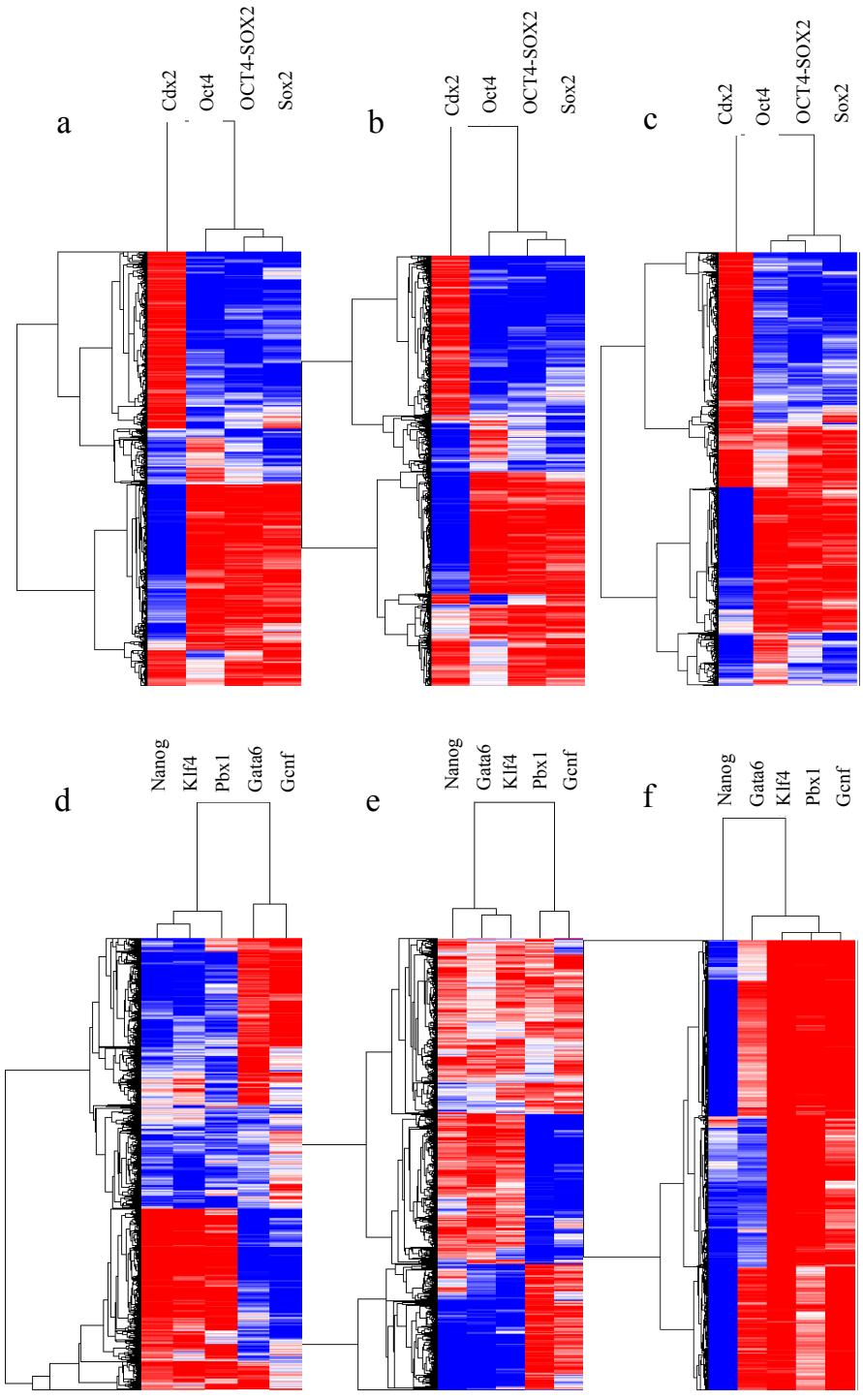


Figure S3.14 - Behaviors of the sub-circuits *Oct4/Cdx2* and *Nanog/Gata6*.

According to Figure 4.3, there exist two sub-circuits (top layer: *Oct4/Cdx2* and bottom layer: *Nanog/Gata6*) to form a hierarchical structure. **(a)** and **(d)**: The RACIPE gene

expression data of the full circuit are split into two sets, each of which corresponds to the genes from either the top sub-circuit (panel a) or the bottom sub-circuit (panel d). The hierarchical clustering results of them are shown in (a) and (f). **(b)**, **(c)**, **(e)** and **(f)**: RACIPE was applied to each sub-circuit whose circuit parameters were randomized with the same ranges as those for the full circuit. The panels show the hierarchical clustering results for each RACIPE dataset. Panels b and e: for each gene, data are normalized by the mean and standard deviation of the expression levels for each sub-circuit; panels c and f: for each gene, data are normalized by the mean and standard deviation of the expression levels for the full circuit. The results show that the RACIPE results remain the same for the *Oct4/Cdx2* sub-circuit when the regulatory links connecting the two sub-circuits are removed. However, the results for the *Nanog/Gata6* sub-circuit are dramatically different no matter which normalization methods are used.

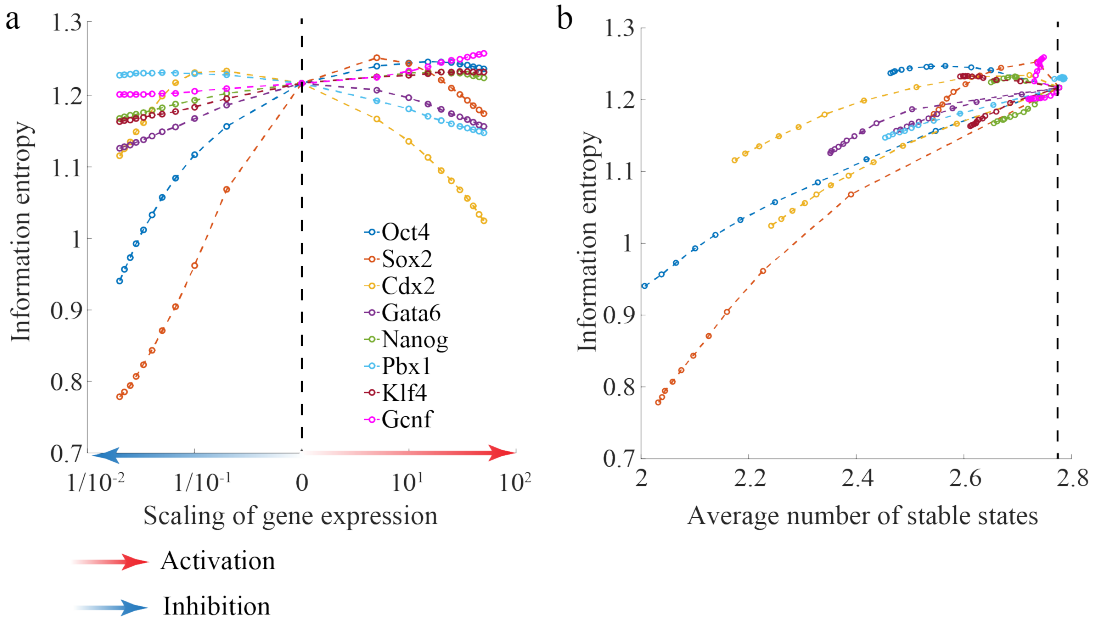


Figure S3.15 - Heterogeneity of the ensemble of the RACIPE models in response to low/high-expression treatments.

(a) Weighted (phylogenetic) information entropy as the function of the level of inhibition/activation for each gene. The directions of inhibition and activation are shown by arrows. **(b)** The information entropy is strongly correlated with the average number of stable states for each RACIPE model. The black dashed vertical lines represent the case without treatment.

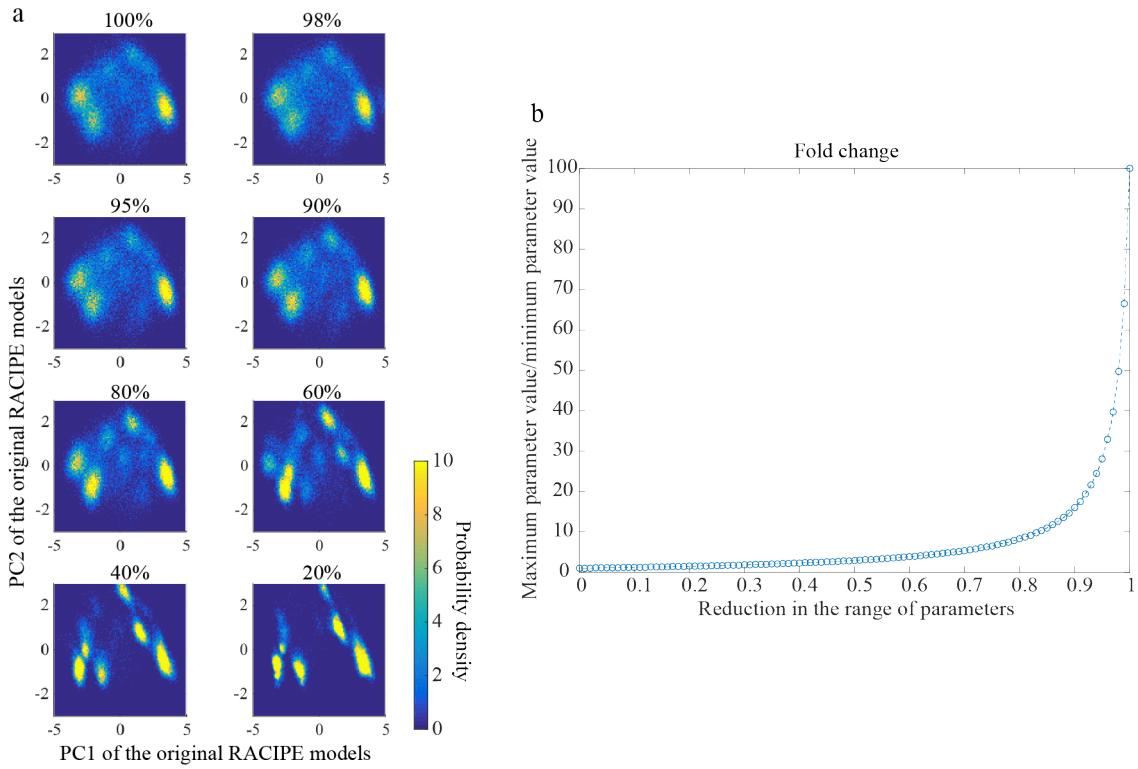


Figure S3.16 - Dependence of the RACIPE gene expression data on the range of the parameters for randomization.

(a) 2D probability density maps of the RACIPE gene expression data projected to the 1st and 2nd principal component axes of the original RACIPE models (as shown in Figure 4.1d). For each small panel, the parameters were randomly selected from different ranges – the original RACIPE protocol was labeled as "100%"; for the other cases, the ranges are reduced to the corresponding percentages. The locations of the major clusters in these plots are similar till the ranges are reduced to 60%. **(b)** The ratio of the maximum and minimum value of fold change (λ) parameters as the function of the shrink of the original range of the parameters for randomization. When there is no reduction (1.0 in the x-axis), the ratio is 100. But the ratio dramatically reduces to about only 10 when the range is reduced to 80%.

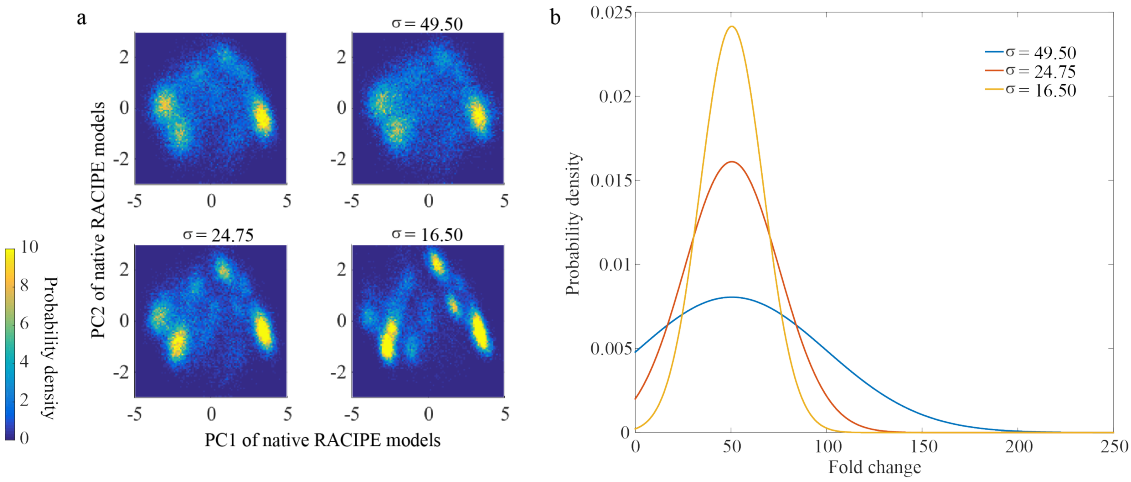


Figure S3.17 - Dependence of the RACIPE gene expression data on the type of the sampling distribution.

So far, we only tested the RACIPE protocol in which the parameters are randomly sampled from uniform distributions. Here we tested RACIPE using Gaussian distribution. **(a)** 2D probability density maps of the RACIPE gene expression data projected to the 1st and 2nd principal component axes of the native RACIPE models (as shown in Figure 4.1d). Top left panel: the standard RACIPE method; the rest panels: the RACIPE protocols using different Gaussian distributions (standard deviations of fold change parameter for each case are shown on the top). The locations of the major clusters in these plots are similar except for the one at bottom right. **(b)** The Gaussian distribution of the fold change (λ) parameters for each case: the blue curve corresponds to the top right panel; the standard deviations of the Gaussian distribution were taken as half of the standard deviations of the uniform distribution, as shown in Table S3.1. The red curve corresponds to the bottom left panel, and the orange curve corresponds to the bottom right panel.

Appendix 4: Modeling putative therapeutic implications of exosome exchange between tumor and immune cells

Table S4.1 - Parameters for the ECI model

Symbol	Value	Unit	Symbol	Value	Unit
Proliferation rate			Apoptosis rate		
g_{D0}	1.35	cell/(μL · day)	k_D	0.03	day^{-1}
g_{C0}	3.6	cell/(μL · day)	k_C	0.03	day^{-1}
g_{K0}	0.06	cell/(μL · day)	k_K	0.03	day^{-1}
Common interactions					
n_{CD1}	2	–	KC_0	450	cell/μL
$CD1_0$	400	cell/μL	λ_{KC}^0	3	–
n_{CD2}	2	–	n_{CC}	4	–
$CD2_0$	50	cell/μL	CC_0	150	cell/μL
λ_{CD2}^+	10	–	λ_{CC}^+	10	–
n_{KD}	2	–	n_{CK1}	2	–
KD_0	150	cell/μL	$CK1_0$	100	cell/μL
λ_{KD}^+	1.5	–	λ_{CK1}^+	3	–
n_{DD}	1	–	n_{CK2}	2	–
DD_0	300	cell/μL	$CK2_0$	400	cell/μL
λ_{DD}^+	2	–	λ_{CK2}^-	0.1	–
n_{DC}	2	–	n_{CK3}	2	–
DC_0	1000	cell/μL	$CK3_0$	400	cell/μL
n_{KC}	2	–	λ_{CK3}^+	3	–
Specific interactions					
Case I					
λ_{CD1}^-	0.1	–	λ_{DC}^0	2	–
Case II					
λ_{CD1}^-	1	–	λ_{DC}^0	1	–
Case III					
λ_{CD1}^-	1	–	λ_{DC}^0	2	–
Case IV					
λ_{CD1}^-	1	–	λ_{DC}^0	1	–
k_{KD}	0.03	day^{-1}			

Noise level γ_x (cells/μL)^2/day $x \in C, D, K$	
$\gamma_x = 312.5$	Fig. 2, Fig. S1, Fig. S2
$\gamma_x = 1250$	Fig. S10
$\gamma_x = 1.125x$	Fig. S11
$\gamma_x = 50$	Fig. 8

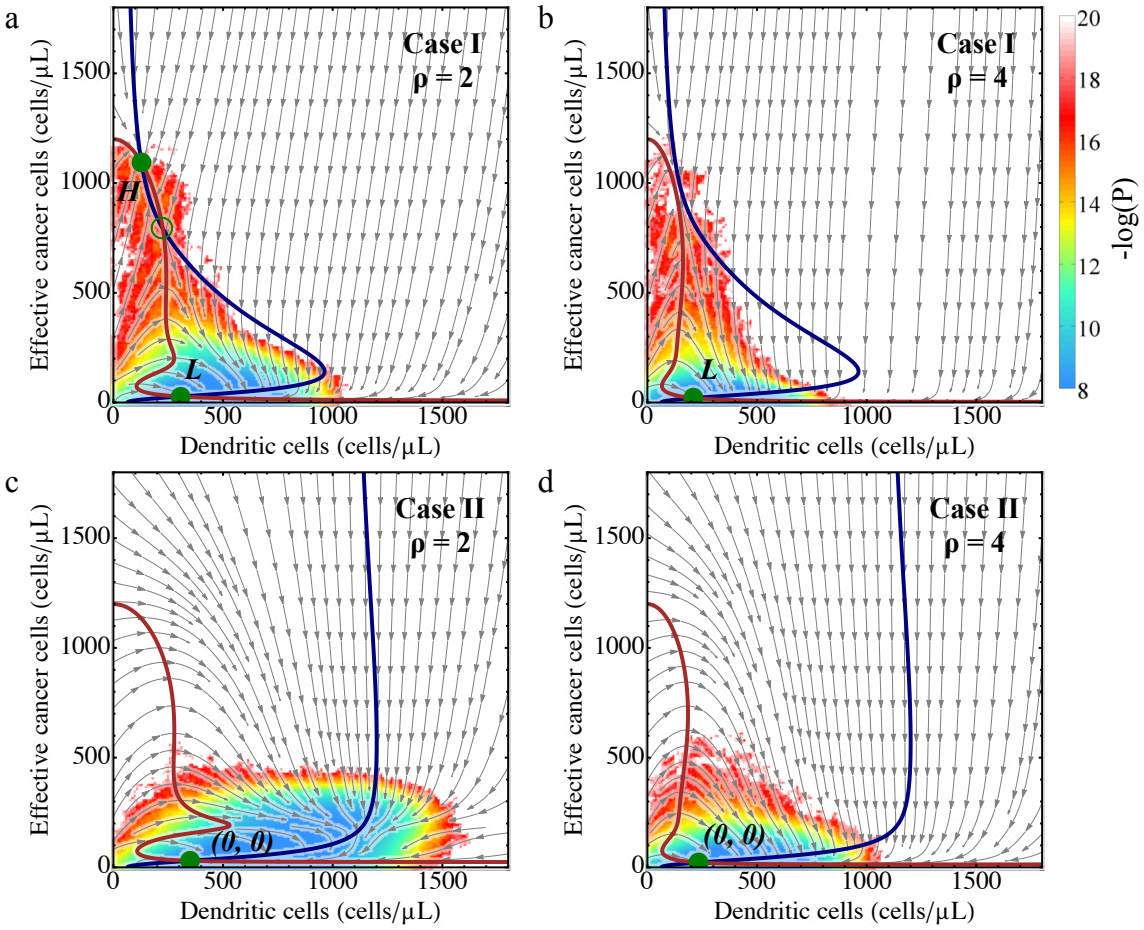


Figure S4.1 - Phase plane analysis for the ECI model.

In each case, a phase plane is constructed by the concentration of the dendritic cells (D, x-axis) and the effective cancer cells (C, y-axis). The nullcline dD/dt = 0 & dK/dt = 0 is shown in solid navy line and dC/dt = 0 & dK/dt = 0 is shown in solid brown line. The intersections of these two nullclines are the steady states, represented by solid green circles for stable states and by hollow green circles for unstable states. The grey arrows represent the vector field in a stream plot. The background colors shows the effective landscape values (-log(P)) computed by stochastic simulations of the network with white Gaussian noise. The states are most probable at the blue regions, and are least probable at the red/white regions. Panel a and b are for case I, and panel c and d are for case II. The immune recognition is high (with value ρ = 2) for a and c, and is higher (with value ρ = 4) for b and d. For both cases, when immune recognition is high enough, the network only has a low cancer load with low immunity level state (L, or (0, 0)) (panel b, c, d). For

case I, even when the immune effect is sufficiently strong, there is still some rare chance to switch to the high cancer (H) state.

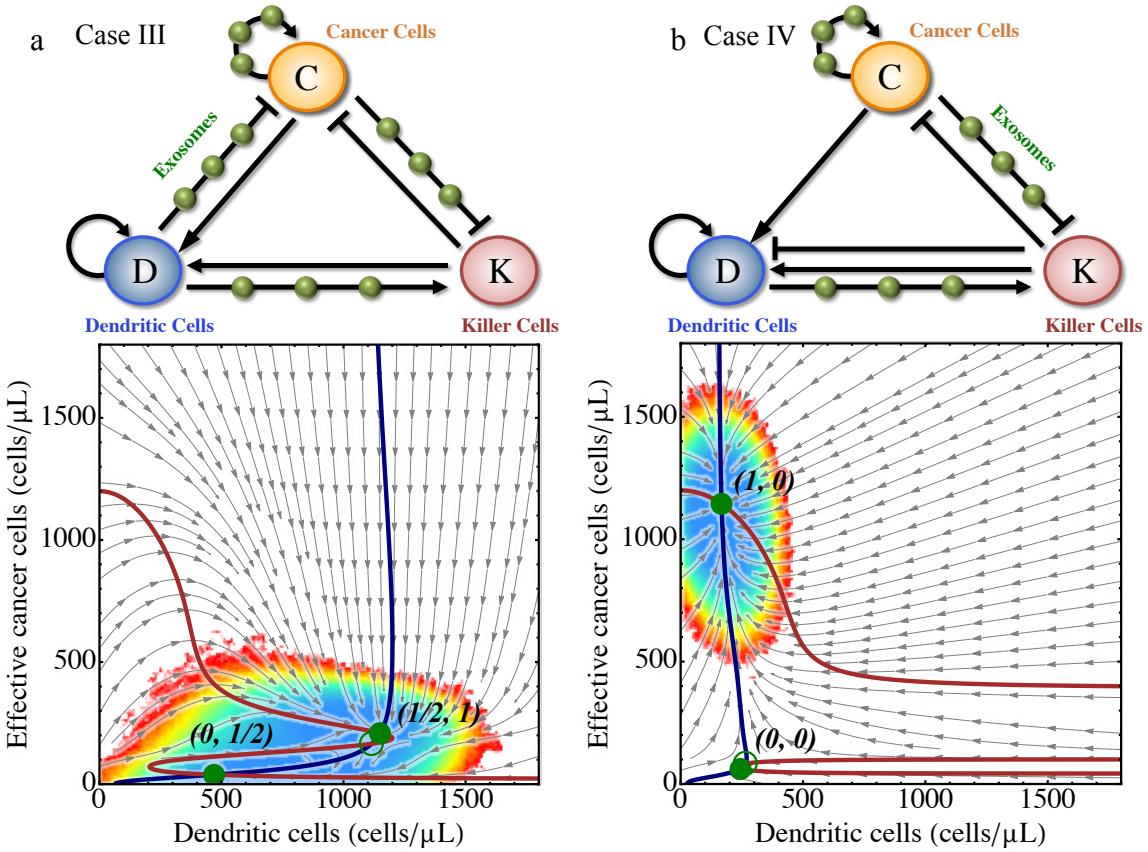


Figure S4.2 - Phase plane analysis for additional ECI models.

The reduced network model contains three major components: the effective cancer cells (C), the dendritic cells (D) and the killer cells (K). The links among different cell types represents the effects of cell-cell communication. An arrow denotes activation; a bar denotes inhibition; the green spheres represent cell communications that are partially mediated by exosomes. Panel a shows the model without exosome-mediated interactions from C to D (only $\lambda_{CD1}^{-}=1$); panel b shows the model without exosome-mediated interactions between C and D ($\lambda_{DC}^{+}=1$ and $\lambda_{CD1}^{-}=1$), but meantime with the apoptosis of matured dendritic cells after they encounter and activate the Naïve T and B cells (with an additional apoptosis term $k_{KD}DH_{CK1}^{S}(C,\lambda_{CK1}^{+})$ added to D's equation (dD/dt)). In all

these cases, we consider ρ = 1, and the rest parameters are kept the same. The corresponding phase plane diagram with effective landscape for each model is shown below the network diagram.

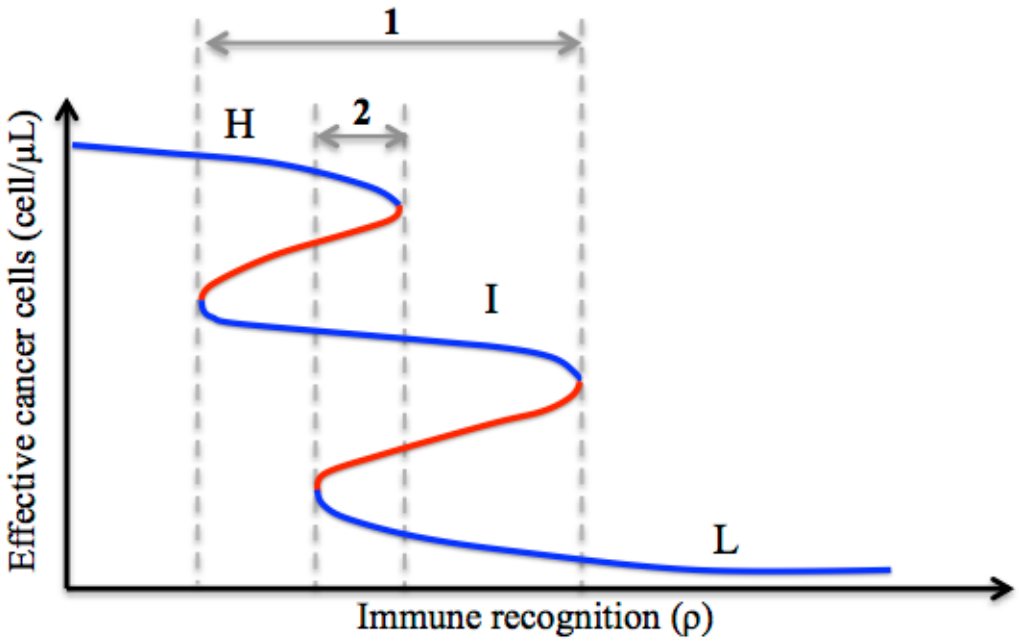


Figure S4.3 - Illustration of the definition of the intermediate state and the tristability.

This illustrative bifurcation diagram is drawn for the steady state concentration of effective cancer cells as the function of the immune recognition ρ. The blue lines represent three segments of stable steady states, while the red lines represent two segments of unstable steady states. We define the existence of intermediate state when there is a total of five such segments. Among these segments, the third (middle) segment (segment labeled as (I)) is defined as the intermediate stable state. The range of existence for the intermediate state in ρ (range 1), is defined as *the range of existence for the intermediate state*. We also define *the range of existence for the tristability* as the range in ρ within which the three stable steady states co-existence (range 2).

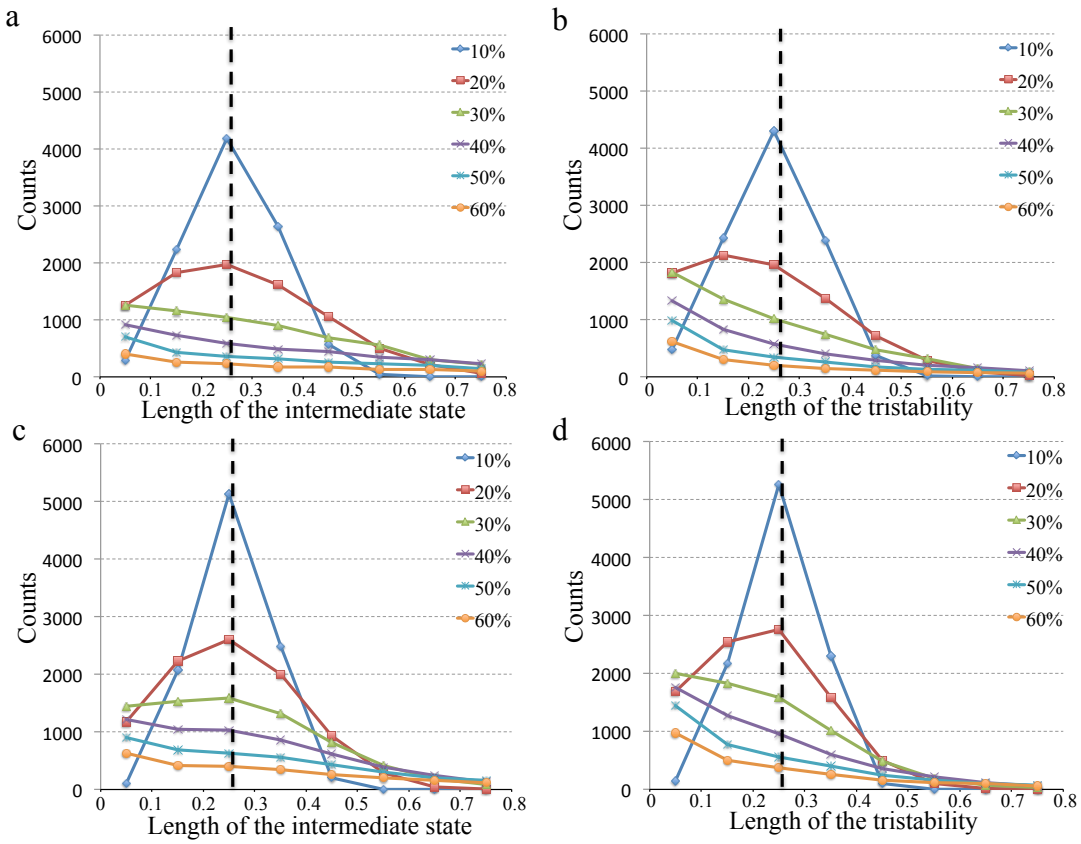
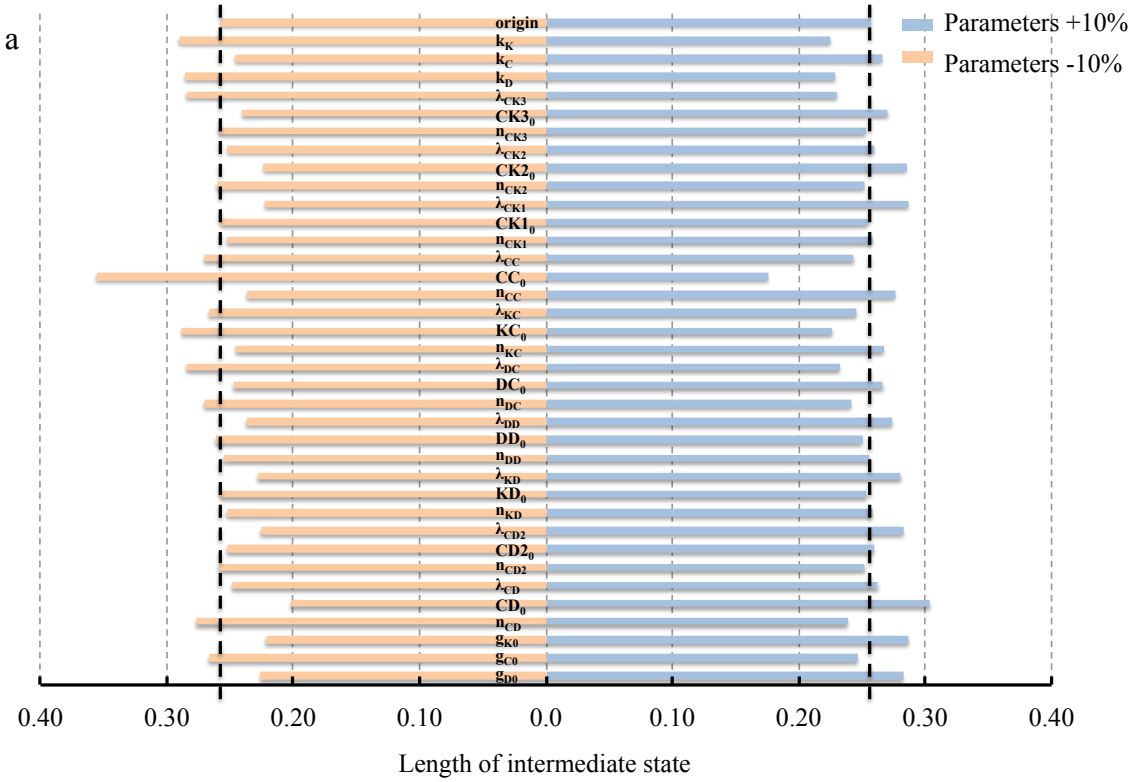
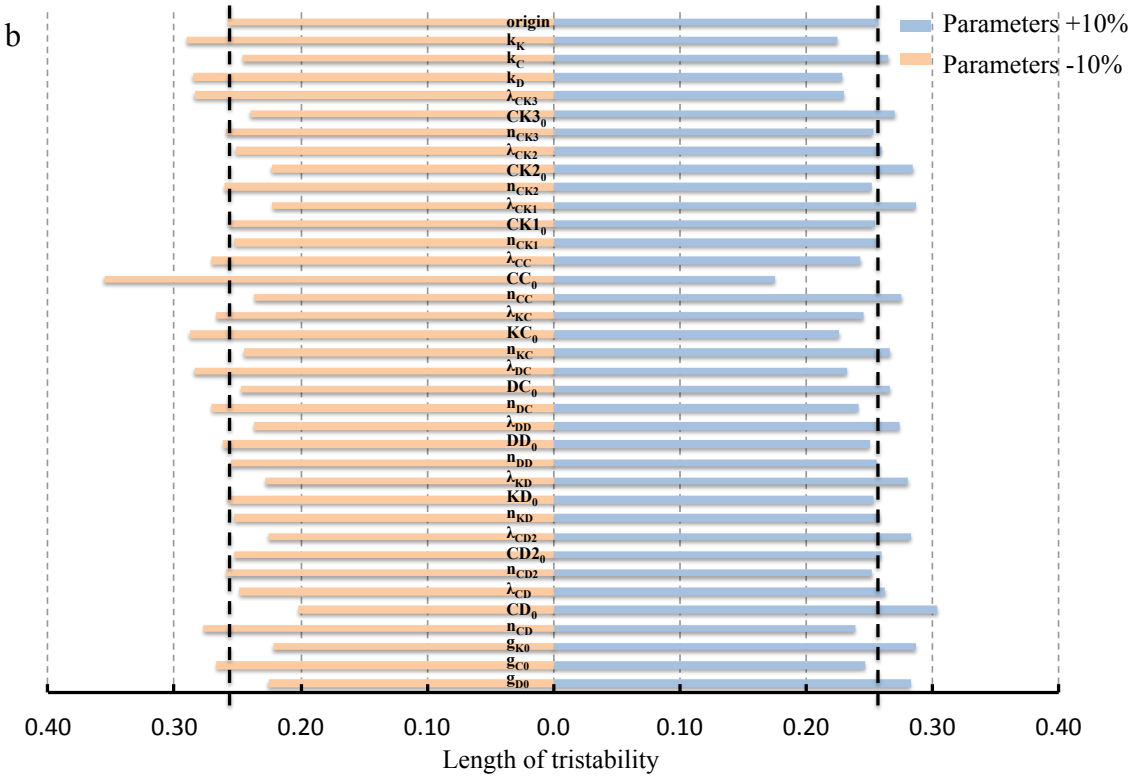
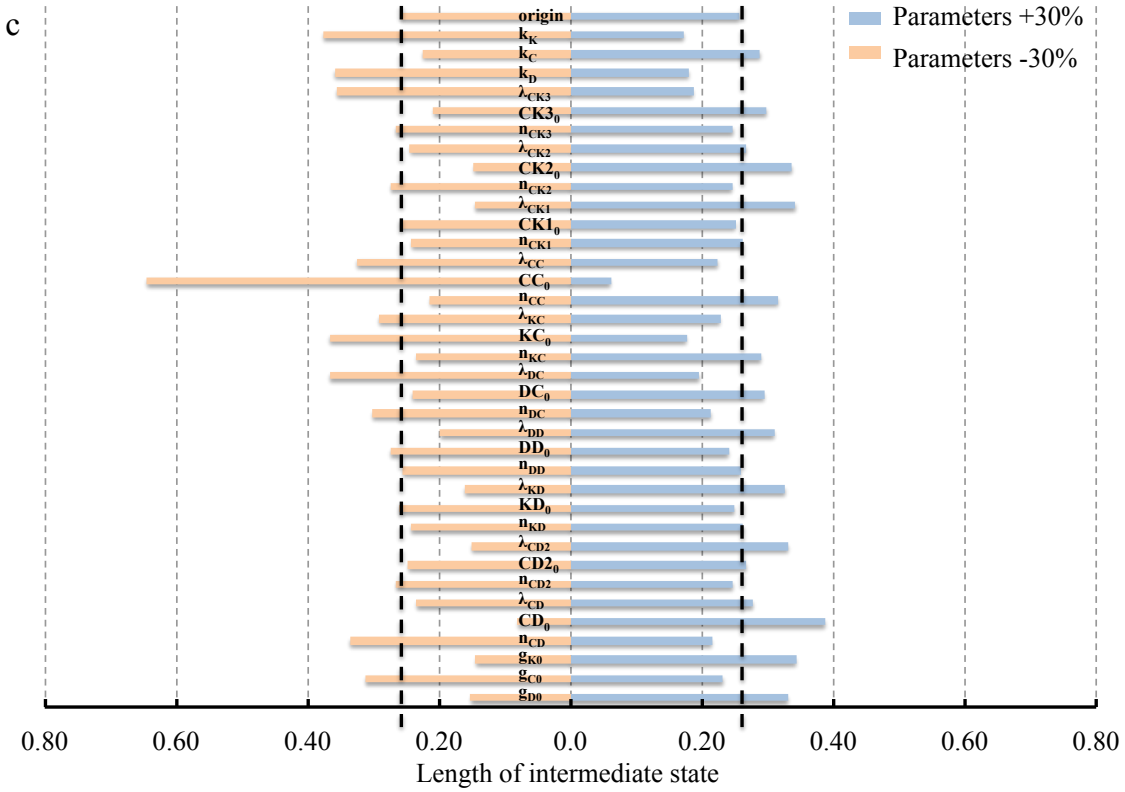


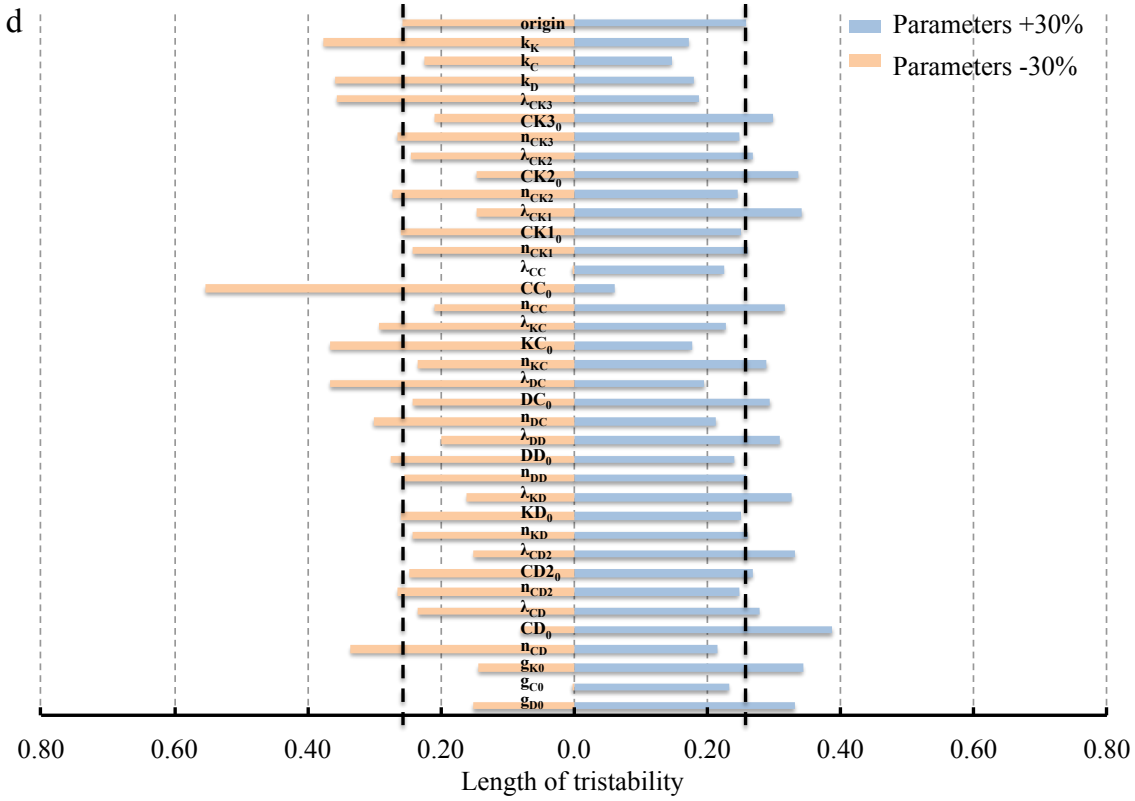
Figure S4.4 - Distribution of the range of existence for the intermediate state and the tristability under different levels of parameter perturbations.

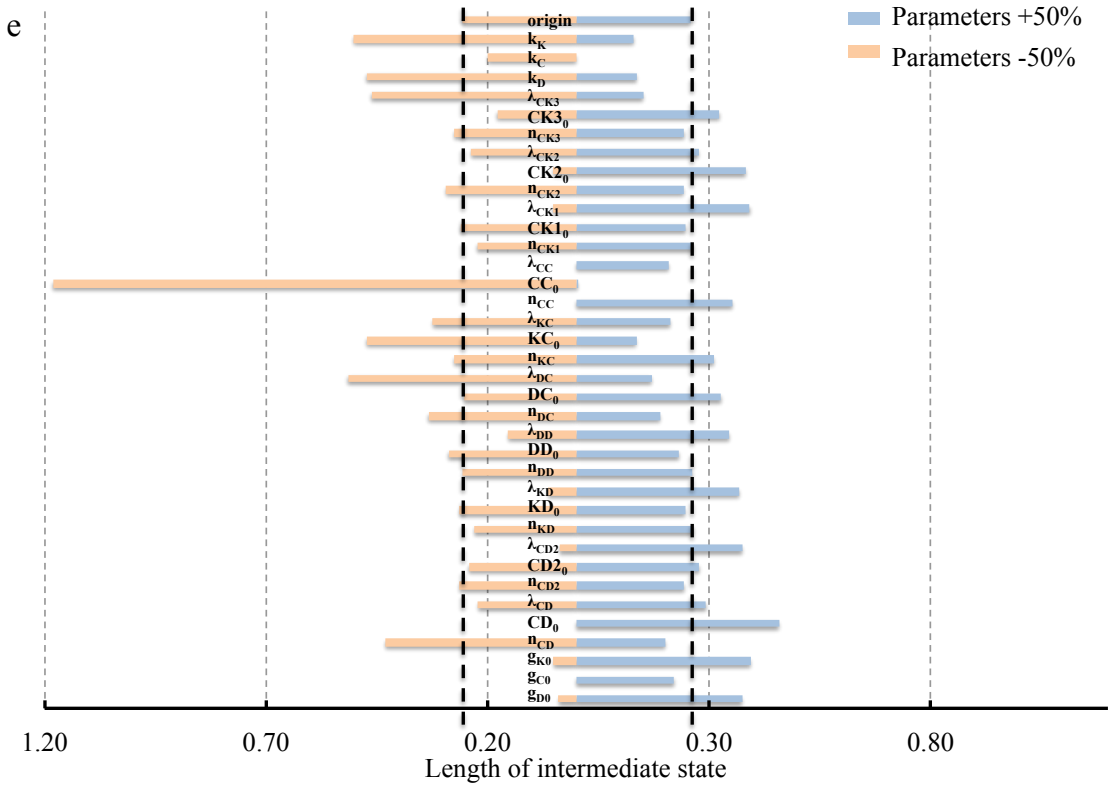
In each test, all 36 parameters were randomly perturbed by a maximum of d% (chosen from 10%, 20% … 60%). Panel a and panel b show the distribution of the range of existence for the intermediate state and the tristability with respect to the immune recognition ρ from a total of 10000 tests, respectively. With the increase of perturbation level, the standard deviations of the average range of existence for the intermediate state and the tristability go up. Panel c and d show the distribution of them when the sensitive parameters (g_{C0} , CC_0 and λ_{CC}^{+}) are excluded from the perturbations in the test. Without these parameters, the standard deviations at same level of perturbation are reduced. The black dashed line shows the baseline, responding to the range of existence for them with the original (unperturbed) parameters.


a
Parameters +10%
Parameters -10%
origin
k_K
k_C
k_D
λ_{CK3}
$CK3_0$
n_{CK3}
λ_{CK2}
$CK2_0$
n_{CK2}
λ_{CK1}
$CK1_0$
n_{CK1}
λ_{CC}
CC_0
n_{CC}
λ_{KC}
KC_0
n_{KC}
λ_{DC}
DC_0
n_{DC}
λ_{DD}
DD_0
n_{DD}
λ_{KD}
KD_0
n_{KD}
λ_{CD2}
$CD2_0$
n_{CD2}
λ_{CD}
CD_0
n_{CD}
g_{K0}
g_{C0}
g_{D0}
0.40 0.30 0.20 0.10 0.0 0.10 0.20 0.30 0.40
Length of intermediate state


b
Parameters +10%
Parameters -10%
origin
k_K
k_C
k_D
λ_{CK3}
$CK3_0$
n_{CK3}
λ_{CK2}
$CK2_0$
n_{CK2}
λ_{CK1}
$CK1_0$
n_{CK1}
λ_{CC}
CC_0
n_{CC}
λ_{KC}
KC_0
n_{KC}
λ_{DC}
DC_0
n_{DC}
λ_{DD}
DD_0
n_{DD}
λ_{KD}
KD_0
n_{KD}
λ_{CD2}
$CD2_0$
n_{CD2}
λ_{CD}
CD_0
n_{CD}
g_{K0}
g_{C0}
g_{D0}
0.40 0.30 0.20 0.10 0.0 0.10 0.20 0.30 0.40
Length of tristability


c
Parameters +30%
Parameters -30%
origin
k_K
k_C
k_D
λ_{CK3}
$CK3_0$
n_{CK3}
λ_{CK2}
$CK2_0$
n_{CK2}
λ_{CK1}
$CK1_0$
n_{CK1}
λ_{CC}
CC_0
n_{CC}
λ_{KC}
KC_0
n_{KC}
λ_{DC}
DC_0
n_{DC}
λ_{DD}
DD_0
n_{DD}
λ_{KD}
KD_0
n_{KD}
λ_{CD2}
$CD2_0$
n_{CD2}
λ_{CD}
CD_0
n_{CD}
g_{K0}
g_{C0}
g_{D0}
0.80
0.60
0.40
0.20
0.0
0.20
0.40
0.60
0.80
Length of intermediate state


d
Parameters +30%
Parameters -30%
origin
k_K
k_C
k_D
λ_{CK3}
$CK3_0$
n_{CK3}
λ_{CK2}
$CK2_0$
n_{CK2}
λ_{CK1}
$CK1_0$
n_{CK1}
λ_{CC}
CC_0
n_{CC}
λ_{KC}
KC_0
n_{KC}
λ_{DC}
DC_0
n_{DC}
λ_{DD}
DD_0
n_{DD}
λ_{KD}
KD_0
n_{KD}
λ_{CD2}
$CD2_0$
n_{CD2}
λ_{CD}
CD_0
n_{CD}
g_{K0}
g_{C0}
g_{D0}
0.80
0.60
0.40
0.20
0.0
0.20
0.40
0.60
0.80
Length of tristability


e
Parameters +50%
Parameters -50%
origin
k_K
k_C
k_D
λ_{CK3}
$CK3_0$
n_{CK3}
λ_{CK2}
$CK2_0$
n_{CK2}
λ_{CK1}
$CK1_0$
n_{CK1}
λ_{CC}
CC_0
n_{CC}
λ_{KC}
KC_0
n_{KC}
λ_{DC}
DC_0
n_{DC}
λ_{DD}
DD_0
n_{DD}
λ_{KD}
KD_0
n_{KD}
λ_{CD2}
$CD2_0$
n_{CD2}
λ_{CD}
CD_0
n_{CD}
g_{K0}
g_{C0}
g_{D0}
1.20
0.70
0.20
0.30
0.80
Length of intermediate state

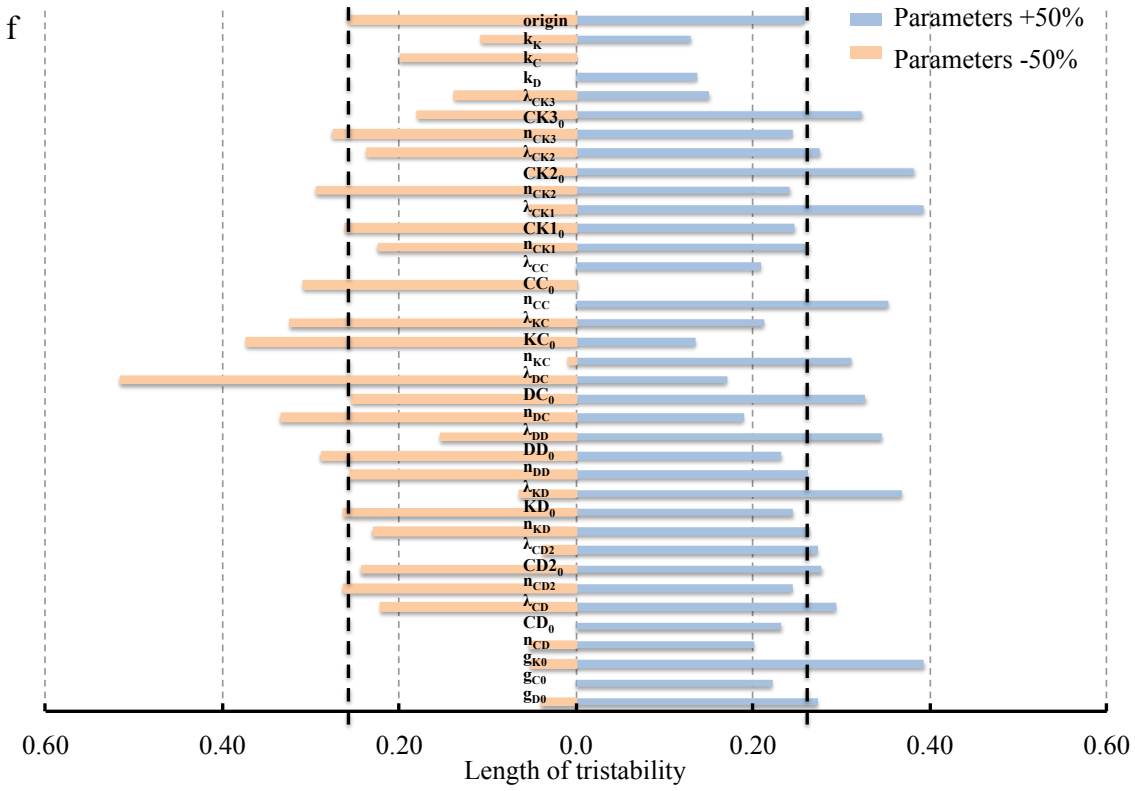
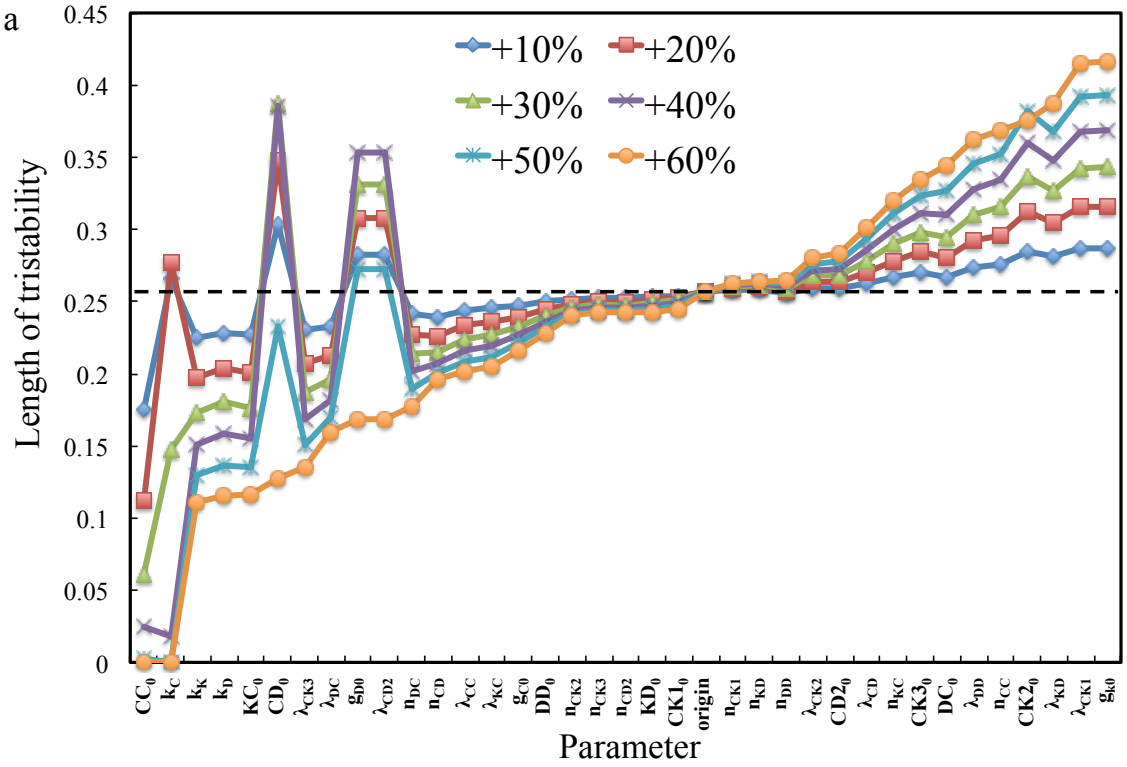

f
Parameters +50%
Parameters -50%
origin
k_K
k_C
k_D
λ_{CK3}
$CK3_0$
n_{CK3}
λ_{CK2}
$CK2_0$
n_{CK2}
λ_{CK1}
$CK1_0$
n_{CK1}
λ_{CC}
CC_0
n_{CC}
λ_{KC}
KC_0
n_{KC}
λ_{DC}
DC_0
n_{DC}
λ_{DD}
DD_0
n_{DD}
λ_{KD}
KD_0
n_{KD}
λ_{CD2}
$CD2_0$
n_{CD2}
λ_{CD}
CD_0
n_{CD}
g_{K0}
g_{C0}
g_{D0}
0.60
0.40
0.20
0.0
0.20
0.40
0.60
Length of tristability

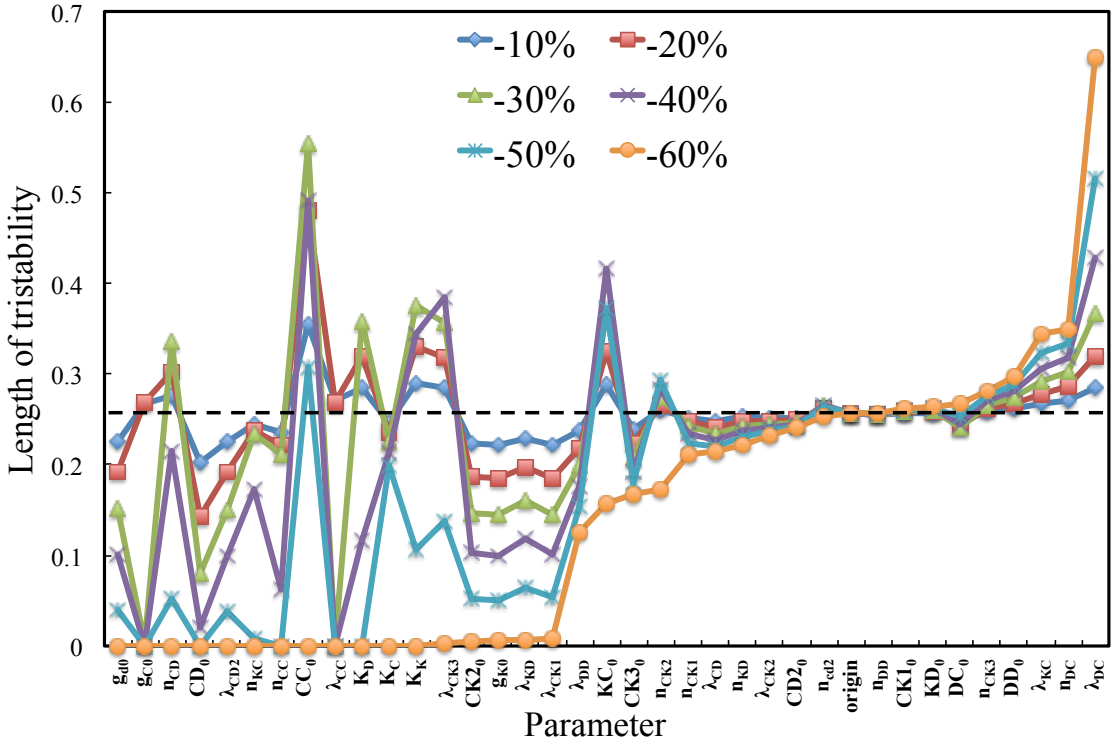
Figure S4.5 - Sensitivity of each model parameter to the range of existence for the intermediate state and the tristability.

In each test, one of 36 parameters is changed by ±10% (panel a and b), ±30% (panel c and d) or ±50% (panel e and f). Then we check the range of existence for the intermediate state and the tristability with respect to the immune recognition ρ. For example, panel a shows the range of existence for the intermediate state for each parameter with -10% (show in orange bars) and +10% (show in blue bars) deviation. The range of existence for the intermediate state for the original parameters is shown at the top and highlighted by vertical dashed black lines. From these results, some sensitive parameters are detected, such as g_{C0}, CC_0 and λ_{CC}^{+}.

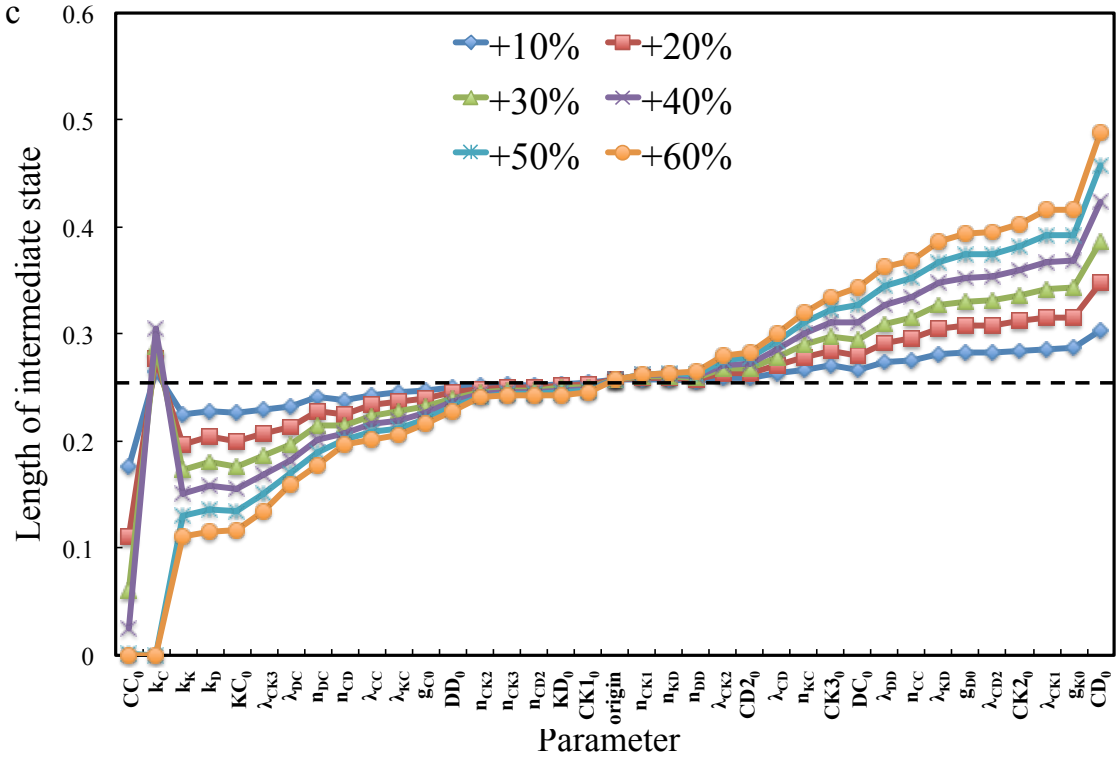
a


+10%
+20%
+30%
+40%
+50%
+60%
Length of tristability
Parameter

b


-10%
-20%
-30%
-40%
-50%
-60%
Length of tristability
Parameter

c


+10%
+20%
+30%
+40%
+50%
+60%
Length of intermediate state
Parameter

d

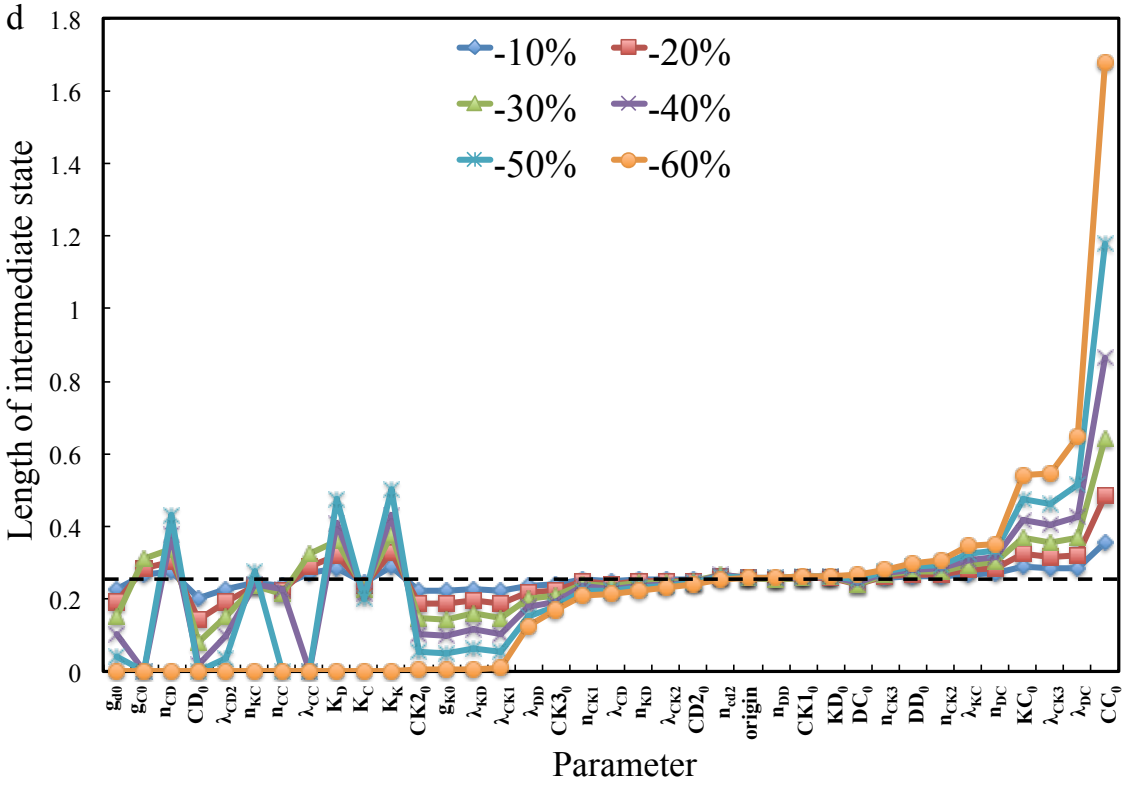

-10%
-20%
-30%
-40%
-50%
-60%
Length of intermediate state
Parameter

Figure S4.6 - The range of existence for the intermediate state and the tristability upon perturbation of each model parameter.

Same as Figure S4.5, one of 36 parameters is changed by ±d% (d is chosen from 10, 20 … 60) each time. The positive deviations of the parameters are shown in panel a and c for the intermediate state and the tristability respectively, while the negative deviations are shown in panel b and d. All the data are plotted in the same order as the one with 60% deviation, which starts from the parameter with the smallest range of existence for the intermediate state or the tristability and end with the one with the largest range.

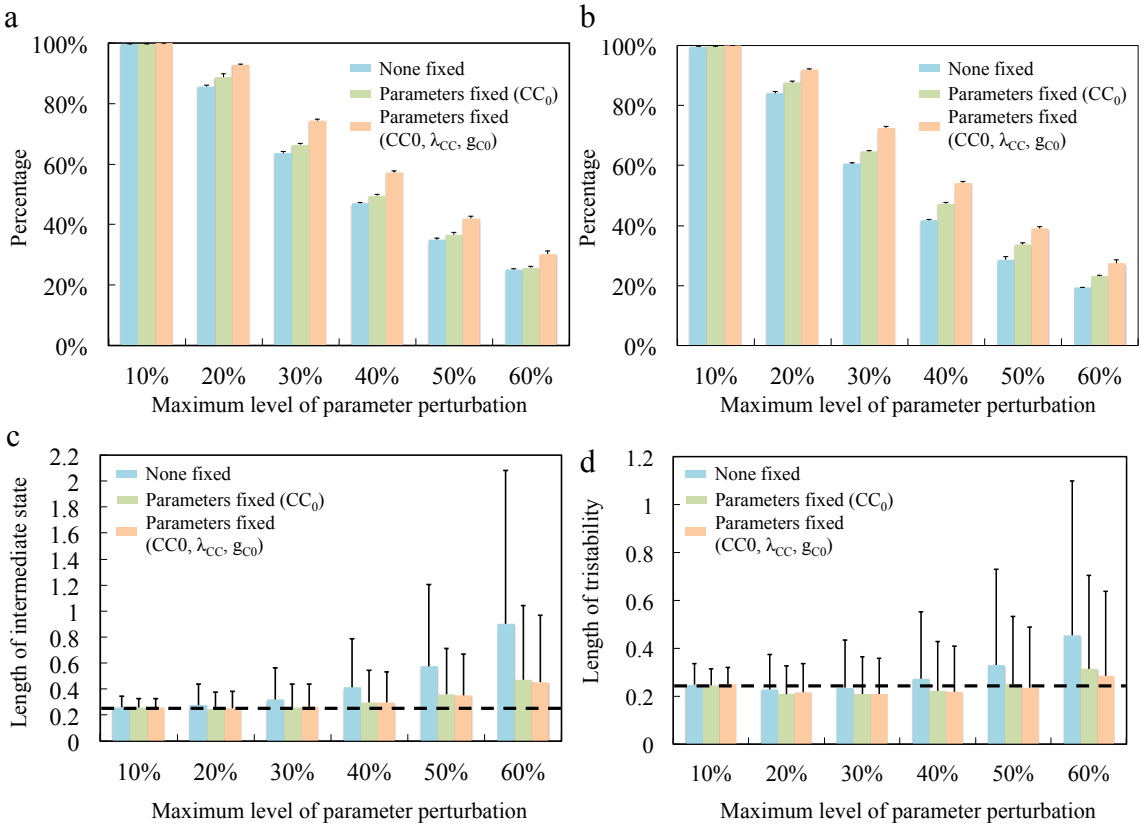


Figure S4.7 - The effects of the sensitive parameters on the existence of the intermediate state and the tristability.

Similar as Figure 5.4, in each test, parameters are randomly perturbed by a maximum of d% (d is chosen from 10, 20 … 60). Then, the bifurcation diagram with respect to the immune recognition ρ is generated to check the existence of the intermediate state and the

tristability. From Figure S4.5 and S4.6 we detect some parameters to which the existence of the intermediate state and the tristability are very sensitive (e.g. g_{C0}, CC_0 and λ_{CC}^{+}). Here, we show the effects of these sensitive parameters on the existence of the intermediate state and the tristability by fixing a part of the sensitive parameters (g_{C0}, CC_0 and λ_{CC}^{+}) in the above test. Panel a and panel b shows the percentage of the existence of the intermediate state and the tristability from a total of 10000 tests, respectively. The light blue column is for the all parameters perturbation, the light green one is for perturbation with CC_0 fixed, and the orange one is for the perturbation with g_{C0}, CC_0 and λ_{CC}^{+} fixed. Panel c and panel d show the mean and standard deviation of the range of existence for the intermediate state and the tristability, respectively. The dotted line shows the baseline, corresponding to the range of existence for them for the case with the unperturbed parameters.

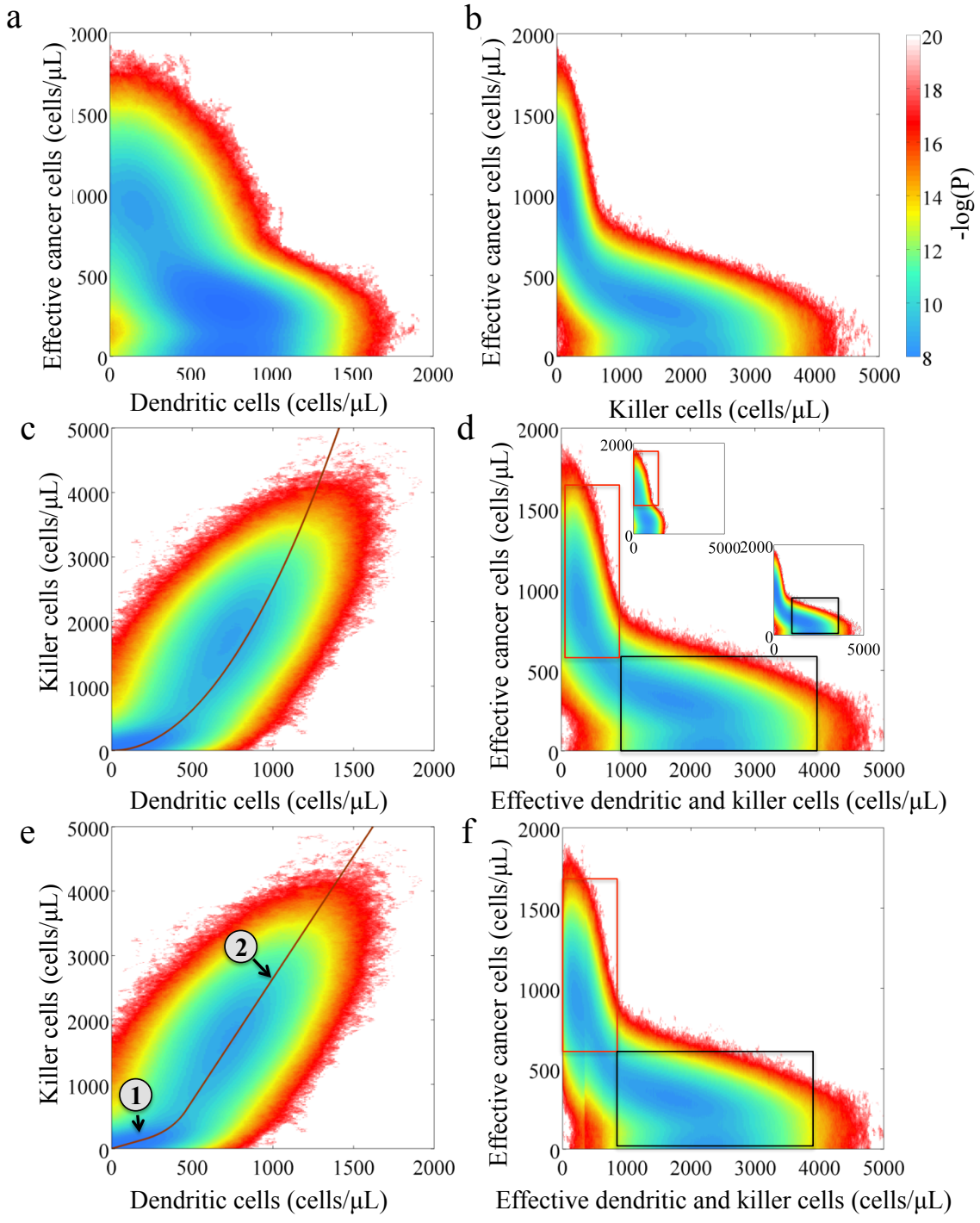


Figure S4.8 - Choice of the reaction coordinates for constructing effective landscapes of the ECI model with white Gaussian noise.

The model has three components (C, D, K), but the effective landscapes, as shown in panel a, b, and c, are usually presented on 2D by averaging the value of the third axis. The effective landscapes computed in this way are very different from each other when different coordinates are selected. Here, for better presentation of the landscape among the three components, we define a new reaction coordinate to describe both D and K. Since K positively correlates to D in a nonlinear way (panel c), we obtain the new reaction coordinate by either fitting the probability distribution to a quadratic function (panel c), or fitting that to two line segments (part 1 and part 2 as shown in panel e) connected by a quadratic Bezier curve (curve between part 1 and 2). We name the new coordinate as "effective dendritic and killer cells". For any state (D, K) in the phase plane DK, we first find the closest state (D_0, K_0) along the new reaction coordinate to the state (D, K), and then define the coordinate value as the arc length of the point (D_0, K_0) from a reference point (origin) along the coordinate. The corresponding effective landscapes are constructed by choosing C as the first axis and the new reaction coordinate as the second axis. The effective landscapes for the two different (yet similar) choices of reaction coordinates, as shown in panel d and f, are similar except that the landscape in panel f is a little discontinuous at the connection part of three segments. In both cases (e and f), the new effective landscape for the high cancer state (red box) resembles to that in panel a (rescaled and shown in the top insert), while that for the intermediate and low cancer state (black box) resembles that in panel b (shown in bottom insert). The plots illustrate the capability of the new reaction coordinate in better describing the effective landscape with reduced dimensions.

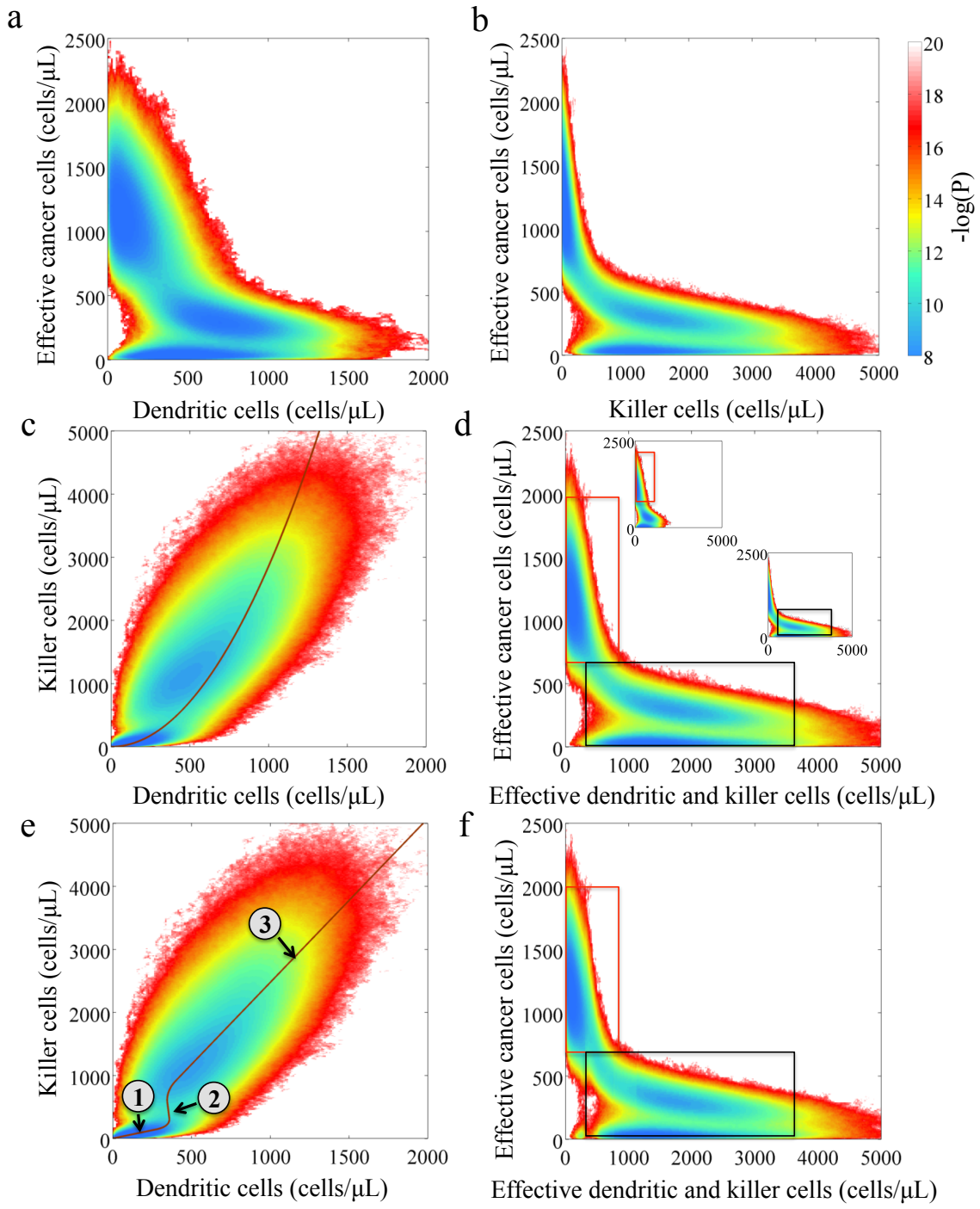


Figure S4.9 - Choice of the reaction coordinates for constructing effective landscapes of the ECI model with shot noise.

Similar to what is shown in Figure S4.8, we obtain the new reaction coordinate by either fitting the probability distribution to a quadratic function (panel c) or to three line segments (part 1, part 2 and part 3 as shown in panel e) connected by two quadratic Bezier curves (between 1, 2 and 2, 3). The effective landscape for high cancer state (red box) resembles that in panel a (rescaled and shown in the top insert), while that for the intermediate and low cancer state (black box) resembles that in panel b (shown in bottom insert).

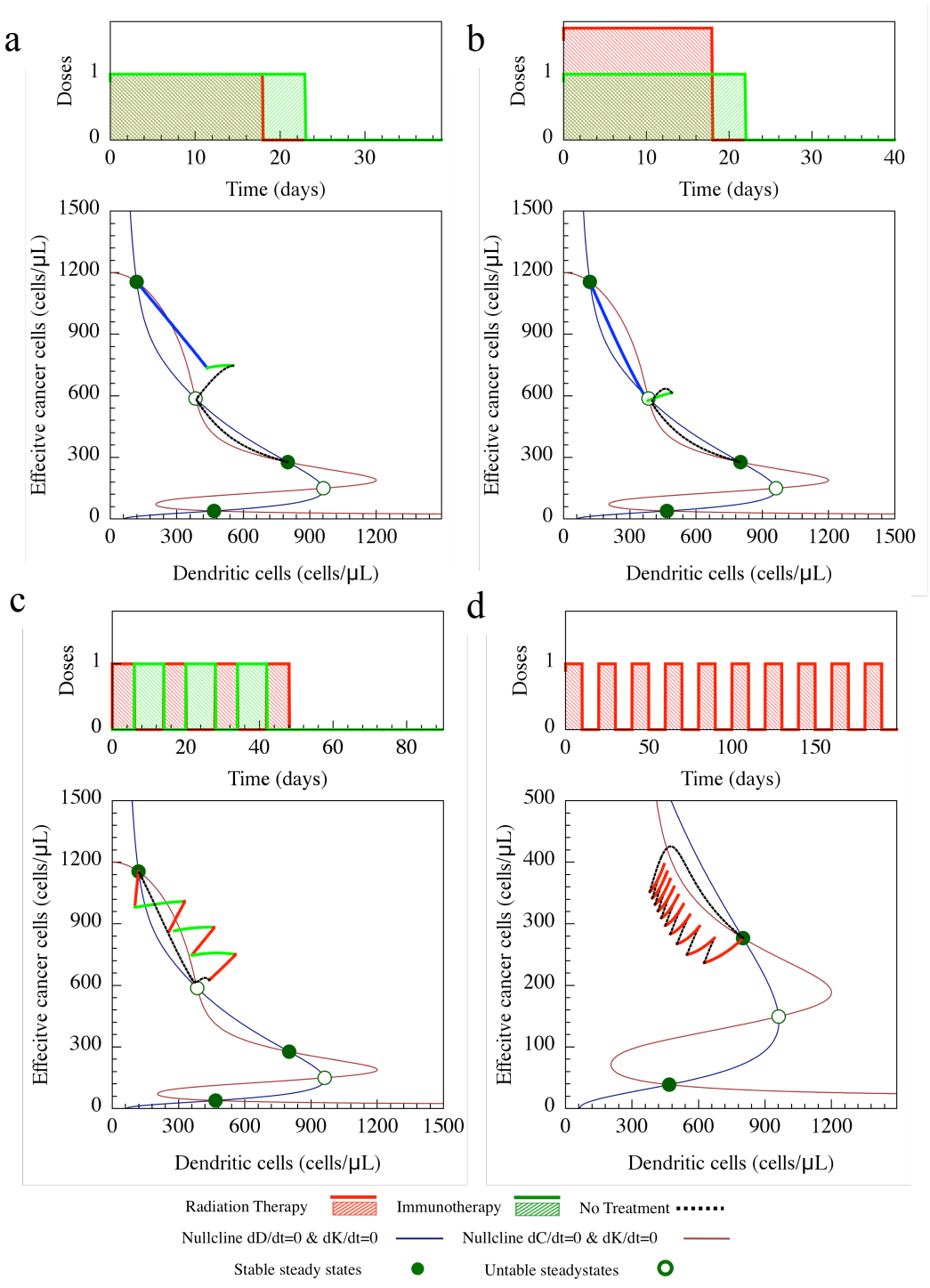


Figure S4.10 - Assessment of therapeutic strategies with the ECI model.

In each case, we show the doses usage (top part) and the dynamical trajectory in the phase plane (bottom part). The session of radiation therapy is shown in red and the

session of DC immunotherapy is shown in green. Panel a shows the treatment from the high cancer state (H) to the intermediate cancer state (I) in case of combined radiation therapy and immunotherapy simultaneously. Panel b shows the similar combined treatment but with stronger radiation effect. Panel c shows alternating combined therapy by both radiation therapy and immunotherapy in same way as right panel of Figure 5.7b, but the total dose usages for each therapy are equal to those for the treatment illustrated in the left panel of Figure 5.7b. Panel d shows the treatment from the intermediate cancer state (I) to the low cancer state (L) by the stand-alone radiation therapy.

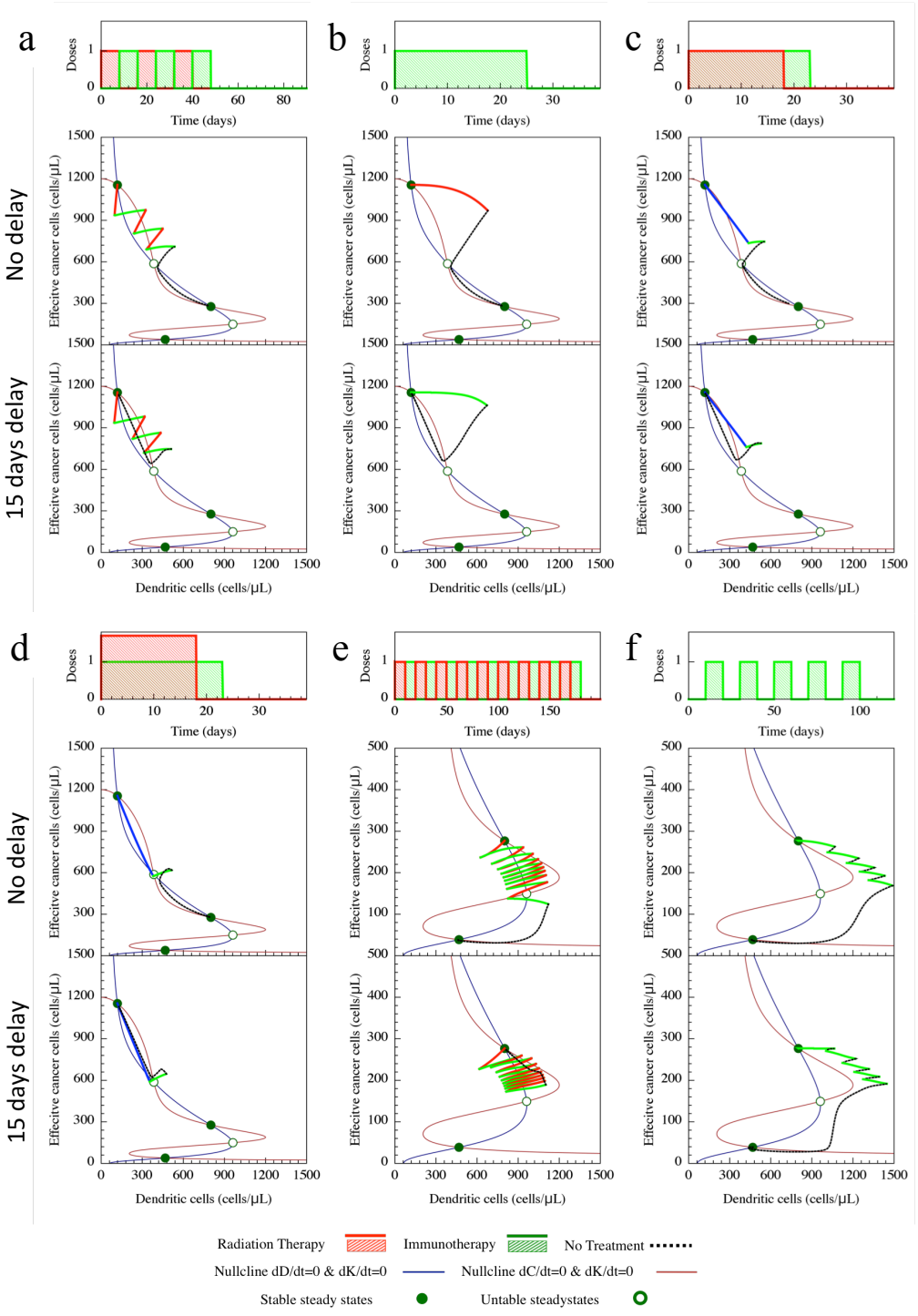


Figure S4.11 - The effects of time delay on the treatments.

In each case, we show the dose usage (top part) and the dynamical trajectory without time delay (middle part). This is the same as some cases in Figure 5.7b and Figure S4.10. We

also show the dynamical trajectory with 15-day delay for the exosome-mediated communications between D and C in the phase plane (bottom part).

References for Appendixes

1. Nimnual AS, Taylor LJ, Bar-Sagi D. Redox-dependent downregulation of Rho by Rac. Nature Cell Biology. 2003;5: 236–241. doi:10.1038/ncb938

2. Arthur WT, Burridge K. RhoA inactivation by p190RhoGAP regulates cell spreading and migration by promoting membrane protrusion and polarity. Molecular Biology of the Cell. 2001;12: 2711–2720.

3. Shen C-H, Chen H-Y, Lin M-S, Li F-Y, Chang C-C, Kuo M-L, et al. Breast tumor kinase phosphorylates p190RhoGAP to regulate rho and ras and promote breast carcinoma growth, migration, and invasion. Cancer research. 2008;68: 7779–7787.

4. Saito K, Ozawa Y, Hibino K, Ohta Y. FilGAP, a Rho/Rho-associated protein kinase–regulated GTPase-activating protein for Rac, controls tumor cell migration. Molecular biology of the cell. 2012;23: 4739–4750.

5. Sanz-Moreno V, Gadea G, Ahn J, Paterson H, Marra P, Pinner S, et al. Rac Activation and Inactivation Control Plasticity of Tumor Cell Movement. Cell. 2008;135: 510–523. doi:10.1016/j.cell.2008.09.043

6. Funato Y, Terabayashi T, Suenaga N. IRSp53/Eps8 Complex Is Important for Positive Regulation of Rac and Cancer Cell Motility/Invasiveness. Cancer Res. 2004;

7. Weiner OD, Rentel MC, Ott A, Brown GE, Jedrychowski M, Yaffe MB, et al. Hem-1 Complexes Are Essential for Rac Activation, Actin Polymerization, and Myosin Regulation during Neutrophil Chemotaxis. PLoS Biology. 2006;4: e38. doi:10.1371/journal.pbio.0040038

8. Kitzing TM, Sahadevan AS, Brandt DT, Knieling H, Hannemann S, Fackler OT, et al. Positive feedback between Dia1, LARG, and RhoA regulates cell morphology and invasion. Genes & Development. 2007;21: 1478–1483. doi:10.1101/gad.424807

9. Chan C-H, Lee S-W, Li C-F, Wang J, Yang W-L, Wu C-Y, et al. Deciphering the transcriptional complex critical for RhoA gene expression and cancer metastasis. Nature Cell Biology. 2010;12: 457–467. doi:10.1038/ncb2047

10. Bustelo XR. A transcriptional cross-talk between RhoA and c-Myc inhibits the RhoA/Rock-dependent cytoskeleton. Landes Bioscience. 2010; doi:10.4161/sgtp.1.1.12986

11. Castellano E, Downward J. RAS Interaction with PI3K: More Than Just Another Effector Pathway. Genes & Cancer. 2011;2: 261–274. doi:10.1177/1947601911408079

12. Lambert JM, Lambert QT, Reuther GW, Malliri A, Siderovski DP, Sondek J, et al. Tiam1 mediates Ras activation of Rac by a PI(3)K-independent mechanism. Nature Cell Biology. 2002; doi:10.1038/ncb833

13. Janes PW, Daly RJ, deFazio A, Sutherland RL. Activation of the Ras signalling pathway in human breast cancer cells overexpressing erbB-2. Oncogene. 1994;9: 3601–3608.

14. Watanabe T, Tsuda M, Makino Y, Ichihara S, Sawa H, Minami A, et al. Adaptor Molecule Crk Is Required for Sustained Phosphorylation of Grb2-Associated Binder 1 and Hepatocyte Growth Factor–Induced Cell Motility of Human Synovial Sarcoma Cell Lines. Mol Cancer Res. 2006;4: 499–510. doi:10.1158/1541-7786.MCR-05-0141

15. Wang D, Paria BC, Zhang Q, Karpurapu M, Li Q, Gerthoffer WT, et al. A Role for Gab1/SHP2 in Thrombin Activation of PAK1 Gene Transfer of Kinase-Dead PAK1 Inhibits Injury-Induced Restenosis. Circulation Research. 2009;104: 1066–1075. doi:10.1161/CIRCRESAHA.109.196691

16. Milo R, Jorgensen P, Moran U, Weber G, Springer M. BioNumbers—the database of key numbers in molecular and cell biology. Nucleic Acids Res. 2010;38: D750–D753. doi:10.1093/nar/gkp889

17. Albert I, Thakar J, Li S, Zhang R, Albert R. Boolean network simulations for life scientists. Source Code for Biology and Medicine. 2008;3: 16. doi:10.1186/1751-0473-3-16

18. Guo G, Huss M, Tong GQ, Wang C, Li Sun L, Clarke ND, et al. Resolution of Cell Fate Decisions Revealed by Single-Cell Gene Expression Analysis from Zygote to Blastocyst. Developmental Cell. 2010;18: 675–685. doi:10.1016/j.devcel.2010.02.012